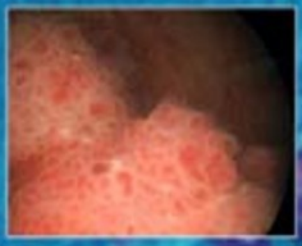
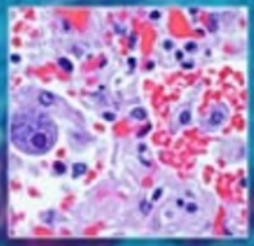
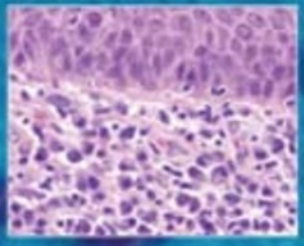
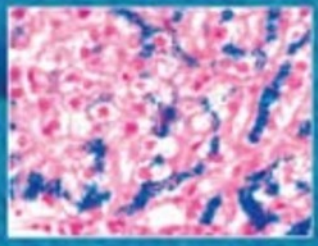
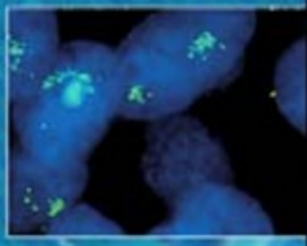
KUMAR
ABBAS
ASTER

Robbins BASIC PATHOLOGY

NINTH EDITION

ELSEVIER
SAUNDERS

Study smart with **Student Consult**

Searchable full text online

Register and activate this title today at studentconsult.com

Activation Code

- **Access the full text online**
- **Download images**
- **Add your own notes and bookmarks**
- **Search across all the Student Consult resources you own online in one place**

ALREADY REGISTERED?

1. Go to studentconsult.com; Sign in
2. Click the "Activate Another Book" button
3. Gently scratch off the surface of the sticker with the edge of a coin to reveal your Pin code
4. Enter it into the "Pin code" box; select the title you've activated from the drop-down menu
5. Click the "Activate Book" button

FIRST-TIME USER?

1. ***REGISTER***
   - Go to studentconsult.com; click "Register Now"
   - Fill in your user information and click "Activate your account"
2. ***ACTIVATE YOUR BOOK***
   - Click the "Activate Another Book" button
   - Gently scratch off the surface of the sticker with the edge of a coin to reveal your Pin code
   - Enter it into the "Pin code" box; select the title you've activated from the drop-down menu
   - Click the "Activate Book" button

For technical assistance: email online.help@elsevier.com
call 800-401-9962 (inside the US) / call +1-314-995-3200 (outside the US)

Robbins Basic Pathology

This page intentionally left blank

ROBBINS

Basic Pathology

NINTH EDITION

Vinay Kumar, MBBS, MD, FRCPath

Donald N. Pritzker Professor
Chair, Department of Pathology
Biologic Sciences Division and
Pritzker School of Medicine
University of Chicago
Chicago, Illinois

Abul K. Abbas, MBBS

Distinguished Professor and Chair
Department of Pathology
University of California San Francisco
San Francisco, California

Jon C. Aster, MD, PhD

Professor of Pathology
Harvard Medical School
Brigham and Women's Hospital
Boston, Massachusetts

ARTIST

James A. Perkins, MS, MFA

ELSEVIER
SAUNDERS

ELSEVIER
SAUNDERS

1600 John F. Kennedy Blvd.
Ste 1800
Philadelphia, PA 19103-2899

ROBBINS BASIC PATHOLOGY 978-1-4377-1781-5
International Edition: 978-0-8089-2432-6

Notices

Knowledge and best practice in this field are constantly changing. As new research and experience broaden our understanding, changes in research methods, professional practices, or medical treatment may become necessary.

Practitioners and researchers must always rely on their own experience and knowledge in evaluating and using any information, methods, compounds, or experiments described herein. In using such information or methods they should be mindful of their own safety and the safety of others, including parties for whom they have a professional responsibility.

With respect to any drug or pharmaceutical products identified, readers are advised to check the most current information provided (i) on procedures featured or (ii) by the manufacturer of each product to be administered, to verify the recommended dose or formula, the method and duration of administration, and contraindications. It is the responsibility of practitioners, relying on their own experience and knowledge of their patients, to make diagnoses, to determine dosages and the best treatment for each individual patient, and to take all appropriate safety precautions.

To the fullest extent of the law, neither the Publisher nor the authors, contributors, or editors, assume any liability for any injury and/or damage to persons or property as a matter of products liability, negligence or otherwise, or from any use or operation of any methods, products, instructions, or ideas contained in the material herein.

Library of Congress Cataloging-in-Publication Data

Robbins basic pathology / [edited by] Vinay Kumar, Abul K. Abbas, Jon C. Aster. – 9th ed.
p. ; cm.
Basic pathology
Includes bibliographical references and index.
ISBN 978-1-4377-1781-5 (hardcover : alk. paper) – ISBN 978-0-8089-2432-6 (International ed. : hardcover : alk. paper)
I. Kumar, Vinay, 1944– II. Abbas, Abul K. III. Aster, Jon C. IV. Robbins, Stanley L. (Stanley Leonard), 1915–2003. V. Title: Basic pathology.
[DNLM: 1. Pathology. QZ 4]
616.07–dc23

2011048699

Executive Content Strategist: William Schmitt
Content Development Manager: Rebecca Gruliow
Publishing Services Manager: Patricia Tannian
Senior Project Manager: Sarah Wunderly
Design Direction: Louis Forgione

Printed in Canada

Last digit is the print number: 9 8 7 6 5 4 3 2 1

DEDICATION

To
Our children and a special grandchild
Kiera Chapman Kumar

This page intentionally left blank

Contributors

Charles E. Alpers, MD
Professor and Vice Chair
Department of Pathology
University of Washington
Seattle, Washington
Kidney and Its Collecting System

Jonathan Epstein, MD
Professor of Pathology, Urology, and Oncology
The Reinhard Professor of Urological Pathology
Director of Surgical Pathology
The Johns Hopkins Medical Institutions
Baltimore, Maryland
Male Genital System and Lower Urinary Tract

Agnes B. Fogo, MD
John L. Shapiro Chair of Pathology
Professor of Pathology, Microbiology, Immunology, Medicine, and Pediatrics
Director, Renal/EM Division of Pathology
Vanderbilt University School of Medicine
Nashville, Tennessee
Kidney and Its Collecting System

Matthew P. Frosch, MD, PhD
Lawrence J. Henderson Associate Professor of Pathology and Health Sciences & Technology
Harvard Medical School
Director, C.S. Kubik Laboratory for Neuropathology
Massachusetts General Hospital
Boston, Massachusetts
Central Nervous System

Aliya Noor Husain, MBBS
Professor
Department of Pathology
The University of Chicago
Chicago, Illinois
Lung

Alexander J.F. Lazar, MD, PhD
Associate Professor
Departments of Pathology and Dermatology
The University of Texas M.D. Anderson Cancer Center
Houston, Texas
Skin

Mark W. Lingen, DDS, PhD
Associate Professor
Department of Pathology
The University of Chicago, Chicago, Illinois
Oral Cavity and Gastrointestinal Tract

Anirban Maitra, MBBS
Professor of Pathology and Oncology
The Johns Hopkins University School of Medicine
Pathologist
The Johns Hopkins Hospital
Baltimore, Maryland
Genetic and Pediatric Diseases; *Pancreas*; *Endocrine System*

Alexander J. McAdam, MD, PhD
Associate Professor of Pathology
Harvard Medical School
Medical Director, Infectious Diseases Diagnostic Laboratory
Children's Hospital
Boston, Massachusetts
General Pathology of Infectious Diseases

Richard N. Mitchell, MD, PhD
Lawrence J. Henderson Professor of Pathology and Health Sciences & Technology
Department of Pathology
Harvard Medical School
Staff Pathologist
Brigham and Women's Hospital
Boston, Massachusetts
Hemodynamic Disorders, Thromboembolism, and Shock; *Blood Vessels*; *Heart*

Peter Pytel, MD
Assistant Professor
Department of Pathology
The University of Chicago
Chicago, Illinois
Peripheral Nerves and Muscles

Andrew E. Rosenberg, MD
Clinical Professor of Pathology
Director, Bone and Soft Tissue Pathology
Department of Pathology
Miller School of Medicine
University of Miami
Miami, Florida
Bones, Joints, and Soft Tissue Tumors

Husain A. Sattar, MD
Assistant Professor of Pathology
The University of Chicago
Chicago, Illinois
Female Genital System and Breast

Arlene H. Sharpe, MD, PhD
Professor of Microbiology and Immunobiology, and Pathology
Harvard Medical School and Brigham and Women's Hospital
Boston, Massachusetts
General Pathology of Infectious Diseases

Thomas Stricker, MD, PhD
Instructor
Department of Pathology
The University of Chicago
Chicago, Illinois
Neoplasia

Neil D. Theise, MD
Professor
Departments of Pathology and Medicine (Digestive Diseases)
Beth Israel Medical Center of Albert Einstein College of Medicine
New York, New York
Liver, Gallbladder, and Biliary Tract

Jerrold R. Turner, MD, PhD
Sara and Harold Lincoln Thompson Professor
Associate Chair
Department of Pathology
The University of Chicago
Chicago, Illinois
Oral Cavity and Gastrointestinal Tract

Wei-Lien Wang, MD
Assistant Professor of Pathology
Section of Soft Tissue and Dermatopathology
The University of Texas M.D. Anderson Cancer Center
Houston, Texas
Skin

Edward C. Klatt, MD
Professor and Academic Administrator
Department of Pathology
Florida State University College of Medicine
Tallahassee, Florida
Photographic Editor

Raminder Kumar, MBBS, MD
Chicago, Illinois
Clinical Editor for Diseases of the Heart, Lung, Pancreas, Oral Cavity and Gastrointestinal Tract, and Liver

Richard N. Mitchell, MD, PhD
Lawrence J. Henderson Professor of Pathology and Health Sciences & Technology
Department of Pathology
Harvard Medical School
Staff Pathologist
Brigham and Women's Hospital
Boston, Massachusetts
Targeted Therapy (Online) Editor

Preface

FORTY YEARS OF BASIC PATHOLOGY

As we reach the 40th year of the publication of *Robbins Basic Pathology,* it is useful to quote Stanley Robbins from the Preface of the first edition (1971):

"Of books as well as men, it may be observed that fat ones contain thin ones struggling to get out. In a sense, this book bears such a relationship to its more substantial progenitor, *Robbins Pathology*. It arose from an appreciation of the modern medical student's dilemma. As the curriculum has become restructured to place greater emphasis on clinical experience, time for reading is correspondingly curtailed. … In writing this book, rare and esoteric lesions are omitted without apology, and infrequent or trivial ones described only briefly. We felt it important, however, to consider rather fully the major disease entities."

The goals of this edition of "baby Robbins" remain true to this vision of Stanley Robbins.

This is an exciting time for students of medicine because the fundamental mechanisms of disease are being unveiled at a breathtaking pace. Pathology is central to understanding the molecular basis of disease, and we have tried to capture the essence of this new knowledge in the ninth edition of *Robbins Basic Pathology.* We firmly believe that pathology forms the scientific foundation of medicine, and advances in the basic sciences ultimately help us in understanding diseases in the individual patient. Thus, while many of the new discoveries in genomics and personalized medicine are covered in the initial chapters on general pathology, we have strived to include the impact of scientific advances on diseases of organ systems described throughout the text. To emphasize the importance of disease mechanisms in the practice of medicine, we have highlighted sections dealing with pathogenesis. In recent years an understanding of the molecular basis of disease has led to the development of "targeted therapies." These are highlighted in the form of "Targeted Therapy" boxes in the online edition of this book. We hope that this new feature will provide examples of "bench-to-bedside" medicine. Although many of the "breakthroughs" in the laboratory have not yet reached the bedside, we have included them in measured "doses" so that students can begin to experience the excitement that is ahead in their careers.

Realizing that the modern medical student feels inundated in trying to synthesize the essentials with the "state of the art," we have continued the use of Summary boxes designed to provide the students with key "take home" messages. These have been retained at the risk of adding a few additional pages to the book since students have uniformly told us that they find them useful.

Many new pieces of four-color art—schematics, flow charts, and diagrammatic representations of disease—have been added to facilitate the understanding of difficult concepts such as the control of the cell cycle, functions of cancer genes, interactions of HIV with its receptors, and the biochemical basis of apoptotic cell death. More illustrations have been added, bringing the total to more than 1,000. Formatting and color palettes of the tables have been changed for greater clarity.

Despite the extensive changes and revisions, our goals remain essentially unaltered. Although we have entered the genomic era, the time-honored tools of gross and microscopic analysis remain useful and morphologic changes are highlighted for ready reference. The strong emphasis on clinicopathologic correlations is maintained, and wherever understood, the impact of molecular pathology on the practice of medicine is emphasized. We are pleased that all of this was accomplished without any "bulge" in the waistline of the text.

We continue to firmly believe that clarity of writing and proper use of language enhance comprehension and facilitate the learning process. Generations of students have told us that they enjoy reading this book. We hope that this edition will be worthy of and possibly enhance the tradition of its forebears.

Acknowledgments

First and foremost, we wish to thank and acknowledge our long-time friend and colleague Dr. Nelson Fausto for his contributions to the previous edition of this book. We continue to benefit from his writing and editing.

Any large endeavor of this type cannot be completed without the help of many individuals. We thank the contributors of various chapters. Many are veterans of the older sibling of this text, the so-called "Big Robbins," and they are listed in the table of contents. To each of them a special thanks. We are fortunate to continue our collaboration with Jim Perkins, whose illustrations bring abstract ideas to life and clarify difficult concepts, and we welcome Dr. Raminder Kumar who edited several chapters for accuracy and appropriateness of the clinical content.

Our assistants, Valerie Driscoll from Chicago, Ana Narvaez from San Francisco, and Muriel Goutas from Boston, deserve thanks for coordinating the tasks.

Many colleagues have enhanced the text by providing helpful critiques in their areas of interest. These include Dr. Rick Aster, who provided "late-breaking news" in the area of climate change science. Many others offered critiques of various chapters. They include Drs. Tony Chang and Neeraj Jolly at the University of Chicago; Drs. Ryan Gill, Andrew Horvai, Marta Margeta, Arie Perry, and Mike Rosenblum of the University of California at San Francisco; Dr. John Stone from Massachusetts General Hospital, Harvard Medical School; Dr. Diego H. Castrillon at UT Southwestern Medical School; and Dr. Victor J. Thannickal of the University of Alabama at Birmingham. Others have provided us with photographic gems from their personal collections. They are individually acknowledged in the credits for their contribution(s). For any unintended omissions we offer our apologies.

Many at Elsevier deserve recognition for their roles in the production of this book. This text was fortunate to be in the hands of Rebecca Gruliow (Manager, Content Development) who has been our partner for several editions. Others deserving of our thanks are Sarah Wunderly (Senor Project Manager) and Lou Forgione (Senior Book Designer). Bill Schmitt, Executive Content Strategist, continues to be our cheerleader and friend. We are especially grateful to the entire production team for tolerating our sometimes next to "impossible" demands and for putting up with our idiosyncrasies during the periods of extreme exhaustion that afflict all authors who undertake what seems like an endless task. We are thankful to the entire Elsevier team for sharing our passion for excellence.

Ventures such as this exact a heavy toll from the families of the authors. We thank them for their tolerance of our absences, both physical and emotional. We are blessed and strengthened by their unconditional support and love, and for their sharing with us the belief that our efforts are worthwhile and useful. We are especially grateful to our wives Raminder Kumar, Ann Abbas, and Erin Malone, who continue to provide steadfast support.

And finally, Vinay Kumar and Abul Abbas welcome Jon Aster, who cut his teeth on the eighth edition of *Pathologic Basis of Disease*, as a co-author and editor. Our partnership thrives because of a shared vision of excellence in teaching despite differences in opinions and individual styles.

VK
AKA
JCA

Contents

Chapters without author names were written by the editors.

See Targeted Therapy available online at studentconsult.com

CHAPTER 1

Cell Injury, Cell Death, and Adaptations

CHAPTER CONTENTS

INTRODUCTION TO PATHOLOGY

Literally translated, *pathology* is the study (*logos*) of disease (*pathos,* suffering). It involves the investigation of the causes of disease and the associated changes at the levels of cells, tissues, and organs, which in turn give rise to the presenting signs and symptoms of the patient. There are two important terms that students will encounter throughout their study of pathology and medicine:

- *Etiology* is the origin of a disease, including the underlying causes and modifying factors. It is now clear that most common diseases, such as hypertension, diabetes, and cancer, are caused by a combination of inherited genetic susceptibility and various environmental triggers. Understanding the genetic and environmental factors underlying diseases is a major theme of modern medicine.
- *Pathogenesis* refers to the steps in the development of disease. It describes how etiologic factors trigger cellular and molecular changes that give rise to the specific functional and structural abnormalities that characterize the disease. Whereas etiology refers to *why* a disease arises, pathogenesis describes *how* a disease develops.

Defining the etiology and pathogenesis of disease not only is essential for understanding a disease but is also the basis for developing rational treatments. Thus, by explaining the causes and development of disease *pathology provides the scientific foundation for the practice of medicine*.

To render diagnoses and guide therapy in clinical practice, pathologists identify changes in the gross or microscopic appearance (*morphology*) of cells and tissues, and biochemical alterations in body fluids (such as blood and urine). Pathologists also use a variety of morphologic, molecular, microbiologic, and immunologic techniques to define the biochemical, structural, and functional changes that occur in cells, tissues, and organs in response to injury. Traditionally, the discipline is divided into general pathology and systemic pathology; the former focuses on the cellular and tissue alterations caused by pathologic stimuli in most tissues, while the latter examines the reactions and abnormalities of different specialized organs. In this book we first cover the broad principles of general pathology and then progress to specific disease processes in individual organs.

OVERVIEW OF CELLULAR RESPONSES TO STRESS AND NOXIOUS STIMULI

Cells are active participants in their environment, constantly adjusting their structure and function to accommodate changing demands and extracellular stresses. Cells normally maintain a steady state called *homeostasis* in which the intracellular milieu is kept within a fairly narrow range of physiologic parameters. As cells encounter physiologic stresses or pathologic stimuli, they can undergo

adaptation, achieving a new steady state and preserving viability and function. The principal adaptive responses are *hypertrophy, hyperplasia, atrophy,* and *metaplasia*. If the adaptive capability is exceeded or if the external stress is inherently harmful, *cell injury* develops (Fig. 1–1). Within certain limits, injury is *reversible,* and cells return to a stable baseline; however, if the stress is severe, persistent and rapid in onset, it results in *irreversible injury* and death of the affected cells. *Cell death* is one of the most crucial events in the evolution of disease in any tissue or organ. It results from diverse causes, including ischemia (lack of blood flow), infections, toxins, and immune reactions. Cell death also is a normal and essential process in embryogenesis, the development of organs, and the maintenance of homeostasis.

The relationships among normal, adapted, and reversibly and irreversibly injured cells are well illustrated by the responses of the heart to different types of stress (Fig. 1–2). Myocardium subjected to persistent increased load, as in hypertension or with a narrowed (stenotic) valve, adapts by undergoing *hypertrophy*—an increase in the size of the individual cells and ultimately the entire heart—to generate the required higher contractile force. If the increased demand is not relieved, or if the myocardium is subjected to reduced blood flow (*ischemia*) from an occluded coronary artery, the muscle cells may undergo injury. Myocardium may be reversibly injured if the stress is mild or the arterial occlusion is incomplete or sufficiently brief, or it may undergo irreversible injury and cell death (*infarction*) after complete or prolonged occlusion. Also of note, stresses

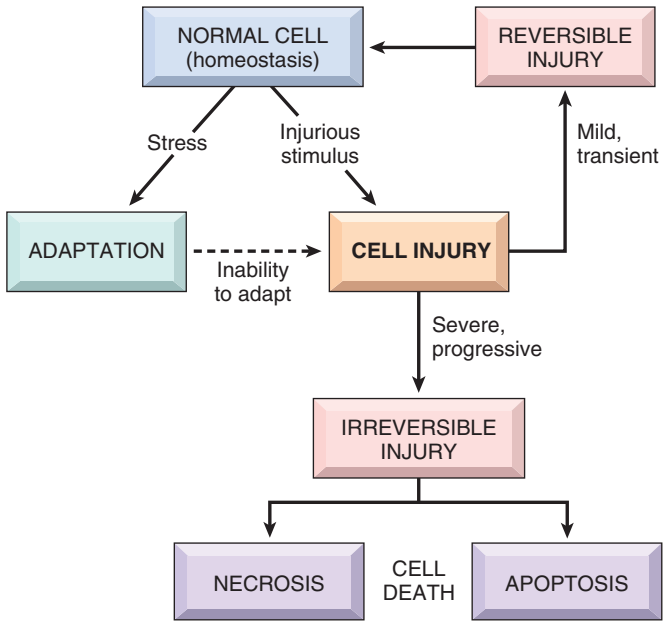


Figure 1–1 Stages in the cellular response to stress and injurious stimuli.

and injury affect not only the morphology but also the functional status of cells and tissues. Thus, reversibly injured myocytes are not dead and may resemble normal myocytes morphologically; however, they are transiently noncontractile, so even mild injury can have a significant

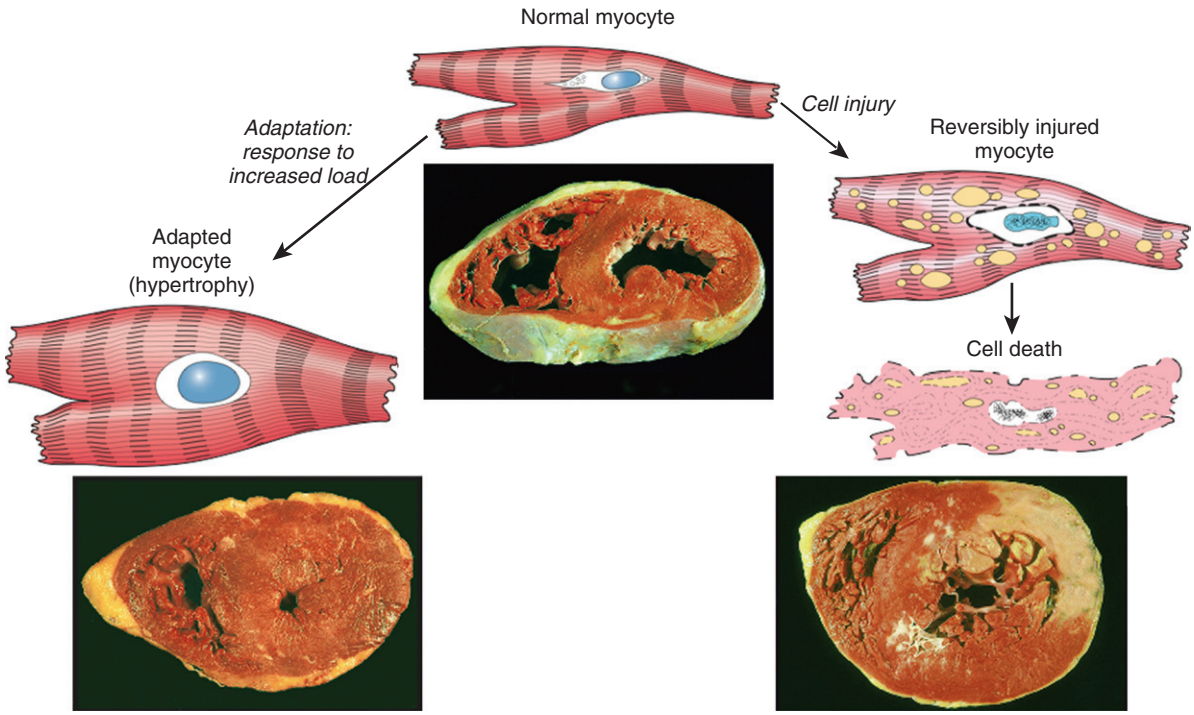


Figure 1–2 The relationship among normal, adapted, reversibly injured, and dead myocardial cells. The cellular adaptation depicted here is hypertrophy, the type of reversible injury is ischemia, and the irreversible injury is ischemic coagulative necrosis. In the example of myocardial hypertrophy (*lower left*), the left ventricular wall is thicker than 2 cm (normal, 1–1.5 cm). Reversibly injured myocardium shows functional effects without any gross or light microscopic changes, or reversible changes like cellular swelling and fatty change (*shown here*). In the specimen showing necrosis (*lower right*) the transmural light area in the posterolateral left ventricle represents an acute myocardial infarction. All three transverse sections of myocardium have been stained with triphenyltetrazolium chloride, an enzyme substrate that colors viable myocardium magenta. Failure to stain is due to enzyme loss after cell death.

clinical impact. Whether a specific form of stress induces adaptation or causes reversible or irreversible injury depends not only on the nature and severity of the stress but also on several other variables, including basal cellular metabolism and blood and nutrient supply.

In this chapter we discuss first how cells adapt to stresses and then the causes, mechanisms, and consequences of the various forms of acute cell damage, including reversible cell injury, subcellular alterations, and cell death. We conclude with three other processes that affect cells and tissues: intracellular accumulations, pathologic calcification, and cell aging.

CELLULAR ADAPTATIONS TO STRESS

Adaptations are reversible changes in the number, size, phenotype, metabolic activity, or functions of cells in response to changes in their environment. *Physiologic adaptations* usually represent responses of cells to normal stimulation by hormones or endogenous chemical mediators (e.g., the hormone-induced enlargement of the breast and uterus during pregnancy). *Pathologic adaptations* are responses to stress that allow cells to modulate their structure and function and thus escape injury. Such adaptations can take several distinct forms.

Hypertrophy

Hypertrophy is an increase in the size of cells resulting in increase in the size of the organ. In contrast, hyperplasia (discussed next) is characterized by an increase in cell number because of proliferation of differentiated cells and replacement by tissue stem cells. Stated another way, in pure hypertrophy there are no new cells, just bigger cells containing increased amounts of structural proteins and organelles. Hyperplasia is an adaptive response in cells capable of replication, whereas hypertrophy occurs when cells have a limited capacity to divide. Hypertrophy and hyperplasia also can occur together, and obviously both result in an enlarged (*hypertrophic*) organ.

Hypertrophy can be physiologic or pathologic and is caused either by increased functional demand or by growth factor or hormonal stimulation.

- The massive physiologic enlargement of the uterus during pregnancy occurs as a consequence of estrogen-stimulated smooth muscle hypertrophy and smooth muscle hyperplasia (Fig. 1–3). In contrast, in response to increased demand the striated muscle cells in both the skeletal muscle and the heart can undergo only hypertrophy because adult muscle cells have a limited capacity to divide. Therefore, the chiseled physique of the avid weightlifter stems solely from the hypertrophy of individual skeletal muscles.
- An example of pathologic cellular hypertrophy is the cardiac enlargement that occurs with hypertension or aortic valve disease (Fig. 1–2).

The mechanisms driving cardiac hypertrophy involve at least two types of signals: *mechanical triggers,* such as stretch, and *trophic triggers,* which typically are soluble mediators that stimulate cell growth, such as growth factors and adrenergic hormones. These stimuli turn on signal transduction pathways that lead to the induction of a number of genes, which in turn stimulate synthesis of many cellular proteins, including growth factors and structural proteins. The result is the synthesis of more proteins and myofilaments per cell, which increases the force generated with each contraction, enabling the cell to meet increased work demands. There may also be a switch of contractile proteins from adult to fetal or neonatal forms. For example, during muscle hypertrophy, the α-myosin heavy chain is replaced by the β form of the myosin heavy chain, which produces slower, more energetically economical contraction.

Whatever the exact mechanisms of hypertrophy, a limit is reached beyond which the enlargement of muscle mass

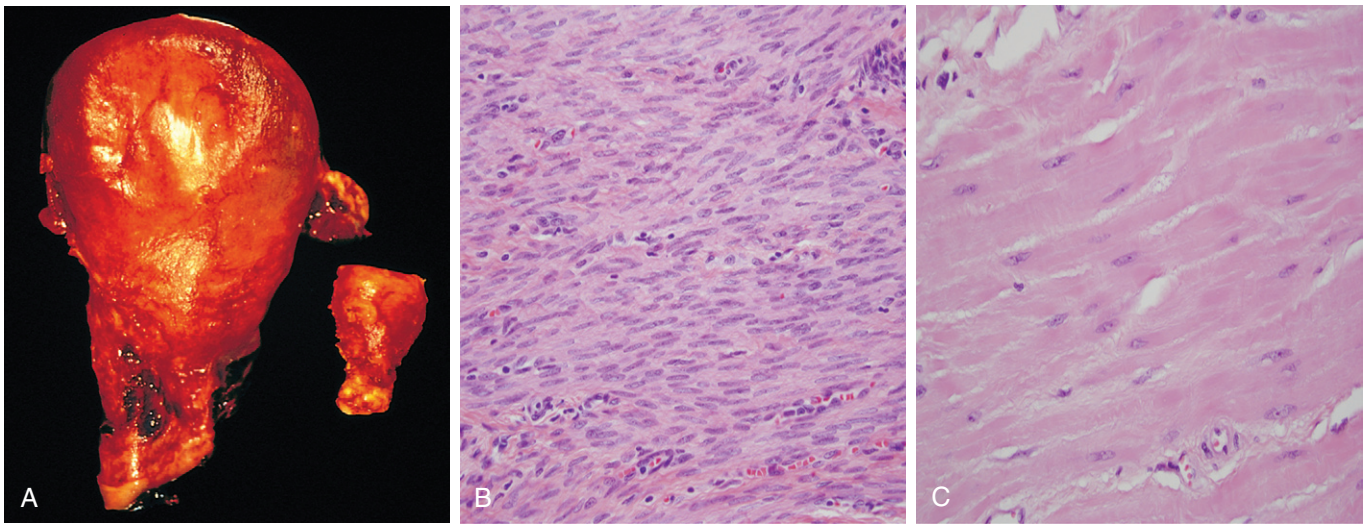


Figure 1–3 Physiologic hypertrophy of the uterus during pregnancy. **A,** Gross appearance of a normal uterus (*right*) and a gravid uterus (*left*) that was removed for postpartum bleeding. **B,** Small spindle-shaped uterine smooth muscle cells from a normal uterus. **C,** Large, plump hypertrophied smooth muscle cells from a gravid uterus; compare with **B.** (**B** and **C,** Same magnification.)

can no longer compensate for the increased burden. When this happens in the heart, several "degenerative" changes occur in the myocardial fibers, of which the most important are fragmentation and loss of myofibrillar contractile elements. The variables that limit continued hypertrophy and cause the regressive changes are incompletely understood. There may be finite limits of the vasculature to adequately supply the enlarged fibers, of the mitochondria to supply adenosine triphosphate (ATP), or of the biosynthetic machinery to provide the contractile proteins or other cytoskeletal elements. The net result of these changes is ventricular dilation and ultimately cardiac failure, a sequence of events that illustrates how *an adaptation to stress can progress to functionally significant cell injury if the stress is not relieved.*

Hyperplasia

As discussed earlier, hyperplasia takes place if the tissue contains cell populations capable of replication; it may occur concurrently with hypertrophy and often in response to the same stimuli.

Hyperplasia can be physiologic or pathologic. In both situations, cellular proliferation is stimulated by growth factors that are produced by a variety of cell types.

- The two types of *physiologic hyperplasia* are (1) *hormonal hyperplasia,* exemplified by the proliferation of the glandular epithelium of the female breast at puberty and during pregnancy, and (2) *compensatory hyperplasia,* in which residual tissue grows after removal or loss of part of an organ. For example, when part of a liver is resected, mitotic activity in the remaining cells begins as early as 12 hours later, eventually restoring the liver to its normal weight. The stimuli for hyperplasia in this setting are polypeptide growth factors produced by uninjured hepatocytes as well as nonparenchymal cells in the liver (Chapter 2). After restoration of the liver mass, cell proliferation is "turned off" by various growth inhibitors.
- Most forms of *pathologic hyperplasia* are caused by excessive hormonal or growth factor stimulation. For example, after a normal menstrual period there is a burst of uterine epithelial proliferation that is normally tightly regulated by stimulation through pituitary hormones and ovarian estrogen and by inhibition through progesterone. However, a disturbed balance between estrogen and progesterone causes endometrial hyperplasia, which is a common cause of abnormal menstrual bleeding. Hyperplasia also is an important response of connective tissue cells in wound healing, in which proliferating fibroblasts and blood vessels aid in repair (Chapter 2). In this process, growth factors are produced by white blood cells (leukocytes) responding to the injury and by cells in the extracellular matrix. Stimulation by growth factors also is involved in the hyperplasia that is associated with certain viral infections; for example, papillomaviruses cause skin warts and mucosal lesions composed of masses of hyperplastic epithelium. Here the growth factors may be encoded by viral genes or by the genes of the infected host cells.

An important point is that in all of these situations, *the hyperplastic process remains controlled; if the signals that initiate it abate, the hyperplasia disappears*. It is this responsiveness to normal regulatory control mechanisms that distinguishes pathologic hyperplasias from cancer, in which the growth control mechanisms become dysregulated or ineffective (Chapter 5). Nevertheless, in many cases, pathologic hyperplasia constitutes a fertile soil in which cancers may eventually arise. For example, patients with hyperplasia of the endometrium are at increased risk of developing endometrial cancer (Chapter 18).

Atrophy

Shrinkage in the size of the cell by the loss of cell substance is known as atrophy. When a sufficient number of cells are involved, the entire tissue or organ diminishes in size, becoming atrophic (Fig. 1–4). Although atrophic cells may have diminished function, they are not dead.

Causes of atrophy include a decreased workload (e.g., immobilization of a limb to permit healing of a fracture),

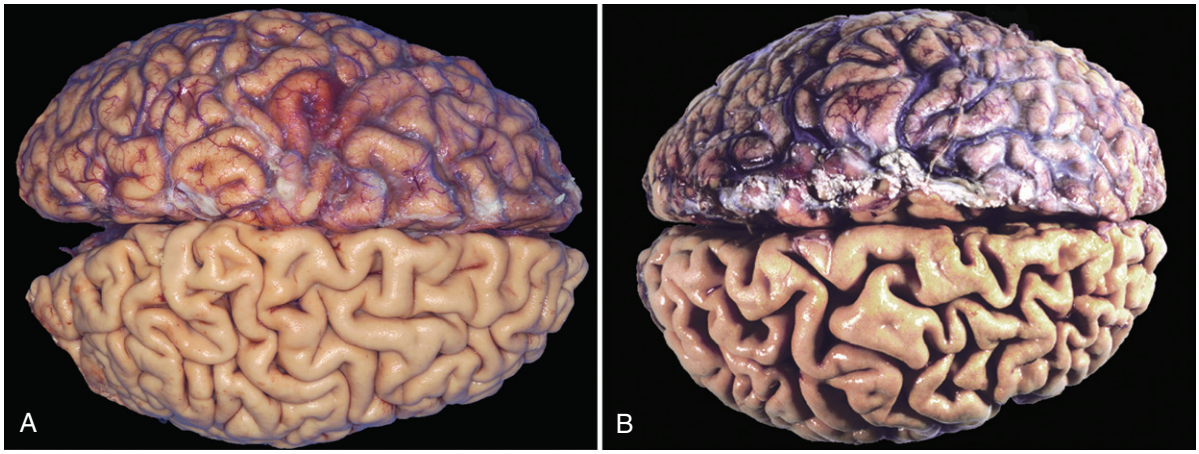


Figure 1–4 Atrophy as seen in the brain. **A,** Normal brain of a young adult. **B,** Atrophy of the brain in an 82-year-old man with atherosclerotic disease. Atrophy of the brain is due to aging and reduced blood supply. Note that loss of brain substance narrows the gyri and widens the sulci. The meninges have been stripped from the bottom half of each specimen to reveal the surface of the brain.

loss of innervation, diminished blood supply, inadequate nutrition, loss of endocrine stimulation, and aging (senile atrophy). Although some of these stimuli are physiologic (e.g., the loss of hormone stimulation in menopause) and others pathologic (e.g., denervation), the fundamental cellular changes are identical. They represent a retreat by the cell to a smaller size at which survival is still possible; a new equilibrium is achieved between cell size and diminished blood supply, nutrition, or trophic stimulation.

The mechanisms of atrophy consist of a combination of decreased protein synthesis and increased protein degradation in cells.

- Protein synthesis decreases because of reduced metabolic activity.
- The degradation of cellular proteins occurs mainly by the *ubiquitin-proteasome pathway.* Nutrient deficiency and disuse may activate ubiquitin ligases, which attach multiple copies of the small peptide ubiquitin to cellular proteins and target them for degradation in proteasomes. This pathway is also thought to be responsible for the accelerated proteolysis seen in a variety of catabolic conditions, including the cachexia associated with cancer.
- In many situations, atrophy is also accompanied by increased *autophagy,* with resulting increases in the number of *autophagic vacuoles.* Autophagy ("self-eating") is the process in which the starved cell eats its own components in an attempt to survive. We describe this process later in the chapter.

Metaplasia

Metaplasia is a reversible change in which one adult cell type (epithelial or mesenchymal) is replaced by another adult cell type. In this type of cellular adaptation, a cell type sensitive to a particular stress is replaced by another cell type better able to withstand the adverse environment. Metaplasia is thought to arise by reprogramming of stem cells to differentiate along a new pathway rather than a phenotypic change (transdifferentiation) of already differentiated cells.

Epithelial metaplasia is exemplified by the squamous change that occurs in the respiratory epithelium of habitual cigarette smokers (Fig. 1–5). The normal ciliated columnar epithelial cells of the trachea and bronchi are focally or widely replaced by stratified squamous epithelial cells. The rugged stratified squamous epithelium may be able to survive the noxious chemicals in cigarette smoke that the more fragile specialized epithelium would not tolerate. *Although the metaplastic squamous epithelium has survival advantages, important protective mechanisms are lost,* such as mucus secretion and ciliary clearance of particulate matter. Epithelial metaplasia is therefore a double-edged sword. Moreover, *the influences that induce metaplastic change, if persistent, may predispose to malignant transformation of the epithelium.* In fact, squamous metaplasia of the respiratory epithelium often coexists with lung cancers composed of malignant squamous cells. It is thought that cigarette smoking initially causes squamous metaplasia, and cancers arise later in some of these altered foci. Since vitamin A is essential for normal epithelial differentiation, its deficiency may also induce squamous metaplasia in the respiratory epithelium. Metaplasia need not always occur in the direction of columnar to squamous epithelium; in chronic gastric reflux, the normal stratified squamous epithelium of the lower esophagus may undergo metaplastic transformation to gastric or intestinal-type columnar epithelium. Metaplasia may also occur in mesenchymal cells but in these situations it is generally a reaction to some pathologic alteration and not an adaptive response to stress. For example, bone is occasionally formed in soft tissues, particularly in foci of injury.

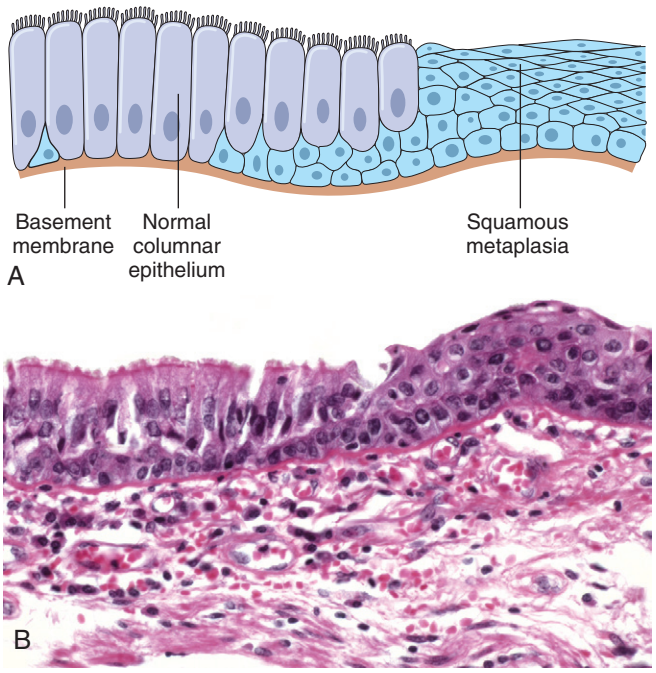


Figure 1–5 Metaplasia of normal columnar (*left*) to squamous epithelium (*right*) in a bronchus, shown schematically (**A**) and histologically (**B**).

SUMMARY

Cellular Adaptations to Stress

- *Hypertrophy:* increased cell and organ size, often in response to increased workload; induced by growth factors produced in response to mechanical stress or other stimuli; occurs in tissues incapable of cell division
- *Hyperplasia:* increased cell numbers in response to hormones and other growth factors; occurs in tissues whose cells are able to divide or contain abundant tissue stem cells
- *Atrophy:* decreased cell and organ size, as a result of decreased nutrient supply or disuse; associated with decreased synthesis of cellular building blocks and increased breakdown of cellular organelles
- *Metaplasia:* change in phenotype of differentiated cells, often in response to chronic irritation, that makes cells better able to withstand the stress; usually induced by altered differentiation pathway of tissue stem cells; may result in reduced functions or increased propensity for malignant transformation

OVERVIEW OF CELL INJURY AND CELL DEATH

As stated at the beginning of the chapter, cell injury results when cells are stressed so severely that they are no longer able to adapt or when cells are exposed to inherently damaging agents or suffer from intrinsic abnormalities (e.g., in DNA or proteins). Different injurious stimuli affect many metabolic pathways and cellular organelles. Injury may progress through a reversible stage and culminate in cell death (Fig. 1–1).

- *Reversible cell injury.* In early stages or mild forms of injury the functional and morphologic changes are reversible if the damaging stimulus is removed. At this stage, although there may be significant structural and functional abnormalities, the injury has typically not progressed to severe membrane damage and nuclear dissolution.
- *Cell death.* With continuing damage, the injury becomes irreversible, at which time the cell cannot recover and it dies. *There are two types of cell death—necrosis and apoptosis—which differ in their mechanisms, morphology, and roles in disease and physiology* (Fig. 1–6 and Table 1–1). When damage to membranes is severe, enzymes leak out of lysosomes, enter the cytoplasm, and digest the cell, resulting in *necrosis.* Cellular contents also leak through the damaged plasma membrane into the extracellular space, where they elicit a host reaction (inflammation). Necrosis is the major pathway of cell death in many commonly encountered injuries, such as those resulting from ischemia, exposure to toxins, various infections, and trauma. When a cell is deprived of growth factors, or the cell's DNA or proteins are damaged beyond repair, typically the cell kills itself by another type of death, called *apoptosis,* which is characterized by nuclear dissolution without complete loss of membrane integrity. *Whereas necrosis is always a pathologic process, apoptosis serves many normal functions and is not necessarily associated with pathologic cell injury. Furthermore, in keeping with its role in certain physiologic processes, apoptosis does not elicit an inflammatory response.* The morphologic features, mechanisms, and significance of these two death pathways are discussed in more detail later in the chapter.

CAUSES OF CELL INJURY

The causes of cell injury range from the gross physical trauma of a motor vehicle accident to the single gene defect that results in a nonfunctional enzyme underlying a

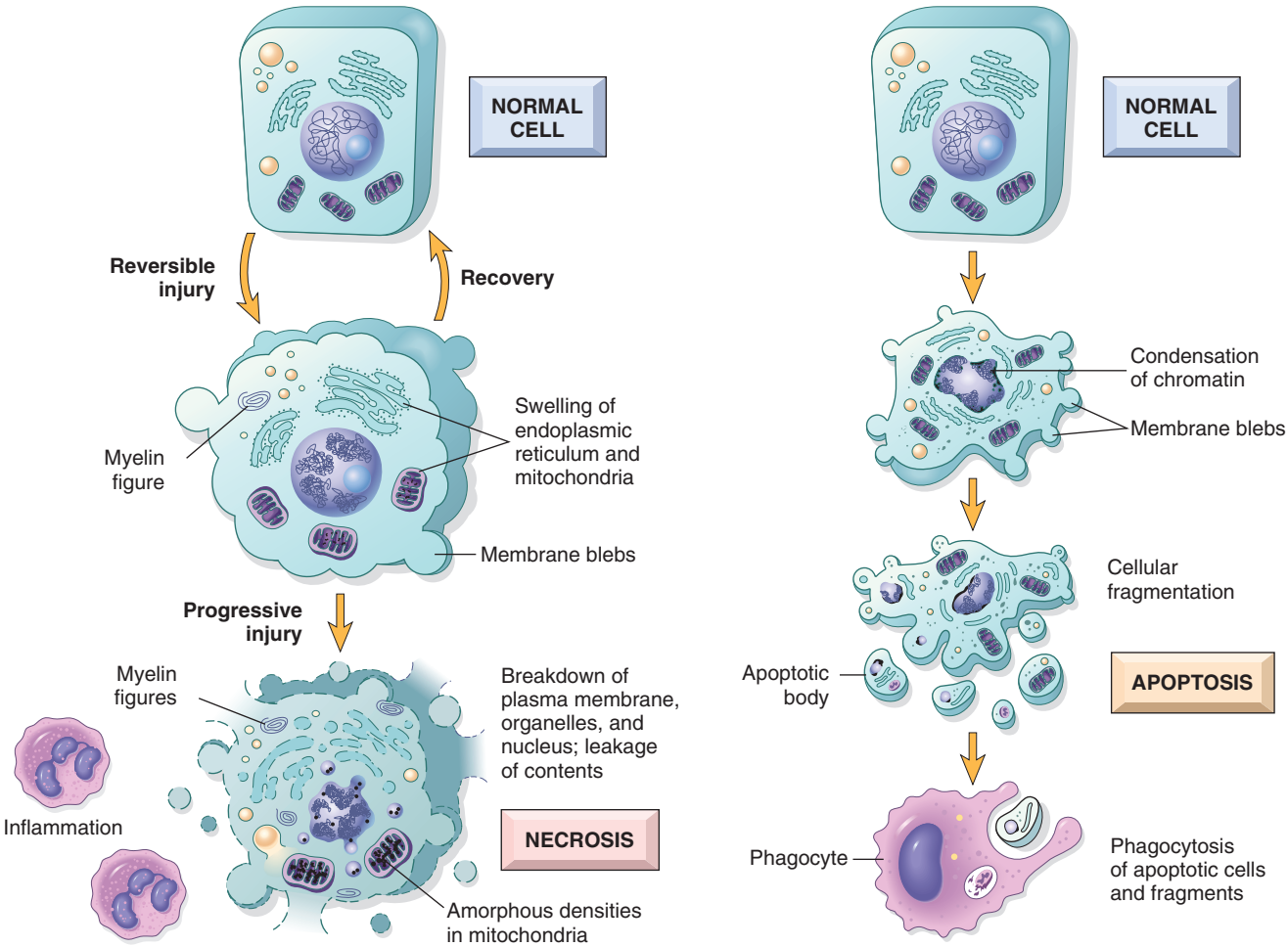


Figure 1–6 Cellular features of necrosis (*left*) and apoptosis (*right*).

Table 1–1 Features of Necrosis and Apoptosis

Feature	Necrosis	Apoptosis
Cell size	Enlarged (swelling)	Reduced (shrinkage)
Nucleus	Pyknosis → karyorrhexis → karyolysis	Fragmentation into nucleosome size fragments
Plasma membrane	Disrupted	Intact; altered structure, especially orientation of lipids
Cellular contents	Enzymatic digestion; may leak out of cell	Intact; may be released in apoptotic bodies
Adjacent inflammation	Frequent	No
Physiologic or pathologic role	Invariably pathologic (culmination of irreversible cell injury)	Often physiologic; means of eliminating unwanted cells; may be pathologic after some forms of cell injury, especially DNA and protein damage

DNA, deoxyribonucleic acid.

specific metabolic disease. Most injurious stimuli can be grouped into the following categories.

Oxygen Deprivation

Hypoxia, or oxygen deficiency, interferes with aerobic oxidative respiration and is an extremely important and common cause of cell injury and death. Hypoxia should be distinguished from *ischemia,* which is a loss of blood supply in a tissue due to impeded arterial flow or reduced venous drainage. While ischemia is the most common cause of hypoxia, oxygen deficiency can also result from inadequate oxygenation of the blood, as in pneumonia, or from reduction in the oxygen-carrying capacity of the blood, as in blood loss anemia or carbon monoxide (CO) poisoning. (CO forms a stable complex with hemoglobin that prevents oxygen binding.)

Chemical Agents

An increasing number of chemical substances that can injure cells are being recognized; even innocuous substances such as glucose, salt, or even water, if absorbed or administered in excess, can so derange the osmotic environment that cell injury or death results. Agents commonly known as poisons cause severe damage at the cellular level by altering membrane permeability, osmotic homeostasis, or the integrity of an enzyme or cofactor, and exposure to such poisons can culminate in the death of the whole organism. Other potentially toxic agents are encountered daily in the environment; these include air pollutants, insecticides, CO, asbestos, and "social stimuli" such as ethanol. Many therapeutic drugs can cause cell or tissue injury in a susceptible patient or if used excessively or inappropriately (Chapter 7). Even oxygen at sufficiently high partial pressures is toxic.

Infectious Agents

Agents of infection range from submicroscopic viruses to meter-long tapeworms; in between are the rickettsiae, bacteria, fungi, and protozoans. The diverse ways in which infectious pathogens cause injury are discussed in Chapter 8.

Immunologic Reactions

Although the immune system defends the body against pathogenic microbes, immune reactions can also result in cell and tissue injury. Examples are autoimmune reactions against one's own tissues and allergic reactions against environmental substances in genetically susceptible individuals (Chapter 4).

Genetic Factors

Genetic aberrations can result in pathologic changes as conspicuous as the congenital malformations associated with Down syndrome or as subtle as the single amino acid substitution in hemoglobin S giving rise to sickle cell anemia (Chapter 6). Genetic defects may cause cell injury as a consequence of deficiency of functional proteins, such as enzymes in inborn errors of metabolism, or accumulation of damaged DNA or misfolded proteins, both of which trigger cell death when they are beyond repair. Genetic variations (polymorphisms) contribute to the development of many complex diseases and can influence the susceptibility of cells to injury by chemicals and other environmental insults.

Nutritional Imbalances

Even in the current era of burgeoning global affluence, nutritional deficiencies remain a major cause of cell injury. Protein–calorie insufficiency among underprivileged populations is only the most obvious example; specific vitamin deficiencies are not uncommon even in developed countries with high standards of living (Chapter 7). Ironically, disorders of nutrition rather than lack of nutrients are also important causes of morbidity and mortality; for example, obesity markedly increases the risk for type 2 diabetes mellitus. Moreover, diets rich in animal fat are strongly implicated in the development of atherosclerosis as well as in increased vulnerability to many disorders, including cancer.

Physical Agents

Trauma, extremes of temperature, radiation, electric shock, and sudden changes in atmospheric pressure all have wide-ranging effects on cells (Chapter 7).

Aging

Cellular senescence leads to alterations in replicative and repair abilities of individual cells and tissues. All of these changes result in a diminished ability to respond to damage and, eventually, the death of cells and of the organism. The mechanisms underlying cellular aging are discussed separately at the end of the chapter.

THE MORPHOLOGY OF CELL AND TISSUE INJURY

It is useful to describe the structural alterations that occur in damaged cells before we discuss the biochemical mechanisms that bring about these changes. All stresses and noxious influences exert their effects first at the molecular or biochemical level. *Cellular function may be lost long before cell death occurs, and the morphologic changes of cell injury (or death) lag far behind both* (Fig. 1–7). For example, myocardial cells become noncontractile after 1 to 2 minutes of ischemia, although they do not die until 20 to 30 minutes of ischemia have elapsed. These myocytes may not appear dead by electron microscopy for 2 to 3 hours, or by light microscopy for 6 to 12 hours.

The cellular derangements of reversible injury can be corrected, and if the injurious stimulus abates, the cell can return to normalcy. Persistent or excessive injury, however, causes cells to pass the nebulous "point of no return" into *irreversible injury* and *cell death.* The events that determine when reversible injury becomes irreversible and progresses to cell death remain poorly understood. The clinical relevance of this question is obvious; if the biochemical and molecular changes that predict cell death can be identified with precision, it may be possible to devise strategies for preventing the transition from reversible to irreversible cell injury. Although there are no definitive morphologic or biochemical correlates of irreversibility, *two phenomena consistently characterize irreversibility: the inability to correct mitochondrial dysfunction* (lack of oxidative phosphorylation and ATP generation) even after resolution of the original injury, and *profound disturbances in membrane function.* As mentioned earlier, injury to lysosomal membranes results in the enzymatic dissolution of the injured cell, which is the culmination of injury progressing to necrosis.

As mentioned earlier, different injurious stimuli may induce death by necrosis or apoptosis (Fig. 1–6 and Table 1–1). Below we describe the morphology of reversible cell injury and necrosis; the sequence of morphologic alterations in these processes is illustrated in Figure 1–6, *left*. Apoptosis has many unique features, and we describe it separately later in the chapter.

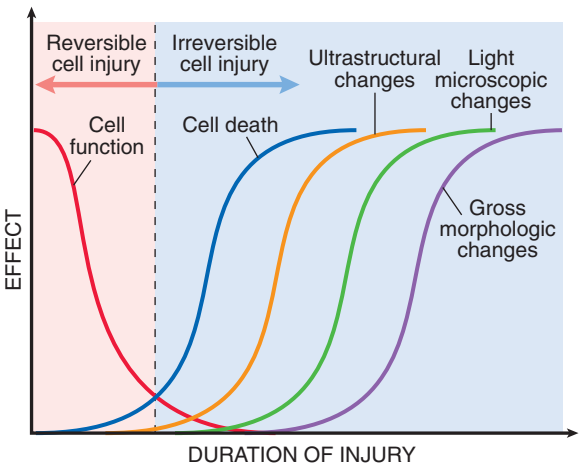


Figure 1–7 The relationship among cellular function, cell death, and the morphologic changes of cell injury. Note that cells may rapidly become nonfunctional after the onset of injury, although they are still viable, with potentially reversible damage; with a longer duration of injury, irreversible injury and cell death may result. Note also that cell death typically precedes ultrastructural, light microscopic, and grossly visible morphologic changes.

Reversible Injury

The two main morphologic correlates of reversible cell injury are *cellular swelling* and *fatty change.* Cellular swelling is the result of failure of energy-dependent ion pumps in the plasma membrane, leading to an inability to maintain ionic and fluid homeostasis. Fatty change occurs in hypoxic injury and in various forms of toxic or metabolic injury and is manifested by the appearance of small or large lipid vacuoles in the cytoplasm. The mechanisms of fatty change are discussed in Chapter 15.

In some situations, potentially injurious insults induce specific alterations in cellular organelles, like the ER. The smooth ER is involved in the metabolism of various chemicals, and cells exposed to these chemicals show hypertrophy of the ER as an adaptive response that may have important functional consequences. For instance, barbiturates are metabolized in the liver by the cytochrome P-450 mixed-function oxidase system found in the smooth ER. Protracted use of barbiturates leads to a state of tolerance, with a decrease in the effects of the drug and the need to use increasing doses. This adaptation is due to increased volume (hypertrophy) of the smooth ER of hepatocytes and consequent increased P-450 enzymatic activity. Although P-450–mediated modification is often thought of as "detoxification," many compounds are rendered *more* injurious by this process; one example is carbon tetrachloride (CCl_4), discussed later. In addition, the products formed by this oxidative metabolism include reactive oxygen species (ROS), which can injure the cell. Cells adapted to one drug have increased capacity to metabolize other compounds handled by the same system. Thus, if patients taking phenobarbital for epilepsy increase their alcohol intake, they may experience a drop in blood concentration of the antiseizure medication to subtherapeutic levels because of induction of smooth ER in response to the alcohol.

MORPHOLOGY

Cellular swelling (Fig. 1–8, *B*), the first manifestation of almost all forms of injury to cells, is a reversible alteration that may be difficult to appreciate with the light microscope, but it may be more apparent at the level of the whole organ. When it affects many cells in an organ, it causes some pallor (as a result of compression of capillaries), increased turgor, and increase in weight of the organ. Microscopic examination may reveal small, clear vacuoles within the cytoplasm; these represent distended and pinched-off segments of the endoplasmic reticulum (ER). This pattern of nonlethal injury is sometimes called **hydropic change** or **vacuolar degeneration. Fatty change** is manifested by the appearance of lipid vacuoles in the cytoplasm. It is principally encountered in cells participating in fat metabolism (e.g., hepatocytes, myocardial cells) and is also reversible. Injured cells may also show increased eosinophilic staining, which becomes much

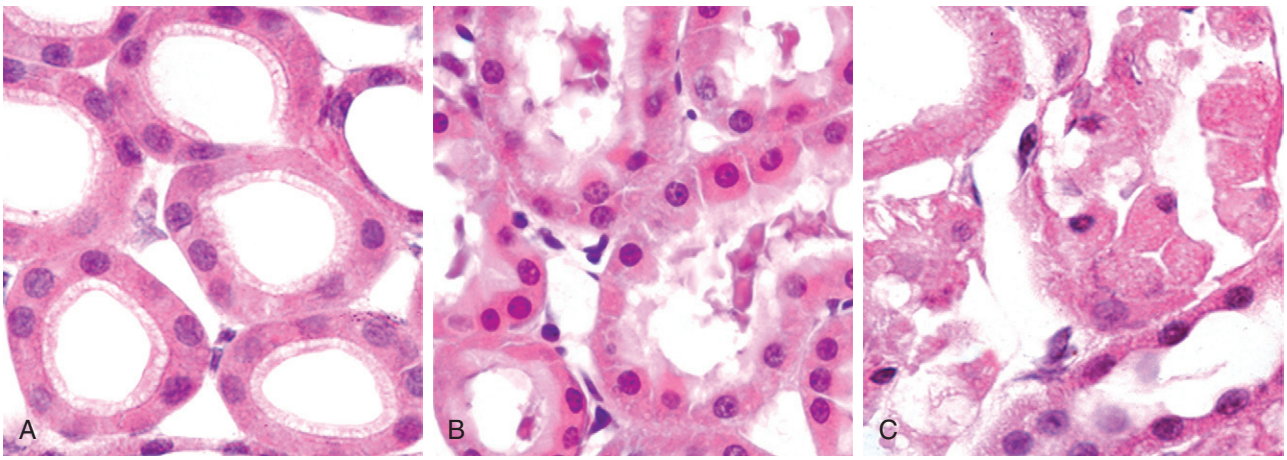


Figure 1–8 Morphologic changes in reversible and irreversible cell injury (necrosis). **A,** Normal kidney tubules with viable epithelial cells. **B,** Early (reversible) ischemic injury showing surface blebs, increased eosinophilia of cytoplasm, and swelling of occasional cells. **C,** Necrotic (irreversible) injury of epithelial cells, with loss of nuclei and fragmentation of cells and leakage of contents.
(Courtesy of Drs. Neal Pinckard and M.A. Venkatachalam, University of Texas Health Sciences Center, San Antonio, Tex.)

more pronounced with progression to necrosis (described further on).

The intracellular changes associated with reversible injury (Fig. 1–6) include (1) plasma membrane alterations such as blebbing, blunting, or distortion of microvilli, and loosening of intercellular attachments; (2) mitochondrial changes such as swelling and the appearance of phospholipid-rich amorphous densities; (3) dilation of the ER with detachment of ribosomes and dissociation of polysomes; and (4) nuclear alterations, with clumping of chromatin. The cytoplasm may contain phospholipid masses, called myelin figures, which are derived from damaged cellular membranes.

Necrosis

Necrosis is the type of cell death that is associated with loss of membrane integrity and leakage of cellular contents culminating in dissolution of cells, largely resulting from the degradative action of enzymes on lethally injured cells. The leaked cellular contents often elicit a local host reaction, called *inflammation,* that attempts to eliminate the dead cells and start the subsequent repair process (Chapter 2). The enzymes responsible for digestion of the cell may be derived from the lysosomes of the dying cells themselves and from the lysosomes of leukocytes that are recruited as part of the inflammatory reaction to the dead cells.

MORPHOLOGY

Necrosis is characterized by changes in the cytoplasm and nuclei of the injured cells (Figs. 1–6, *left,* and 1–8, *C*).

- **Cytoplasmic changes.** Necrotic cells show **increased eosinophilia** (i.e., pink staining from the eosin dye—the E in the hematoxylin and eosin [H&E] stain), attributable in part to increased binding of eosin to denatured cytoplasmic proteins and in part to loss of the basophilia that is normally imparted by the ribonucleic acid (RNA) in the cytoplasm (basophilia is the blue staining from the hematoxylin dye—the H in "H&E"). Compared with viable cells, the cell may have a more glassy, homogeneous appearance, mostly because of the loss of glycogen particles. Myelin figures are more prominent in necrotic cells than during reversible injury. When enzymes have digested cytoplasmic organelles, the cytoplasm becomes vacuolated and appears "moth-eaten." By electron microscopy, necrotic cells are characterized by discontinuities in plasma and organelle membranes, marked dilation of mitochondria with the appearance of large amorphous densities, disruption of lysosomes, and intracytoplasmic myelin figures.
- **Nuclear changes.** Nuclear changes assume one of three patterns, all due to breakdown of DNA and chromatin. The basophilia of the chromatin may fade **(karyolysis),** presumably secondary to deoxyribonuclease (DNase) activity. A second pattern is **pyknosis,** characterized by nuclear shrinkage and increased basophilia; the DNA condenses into a solid shrunken mass. In the third pattern, **karyorrhexis,** the pyknotic nucleus undergoes fragmentation. In 1 to 2 days, the nucleus in a dead cell may completely disappear. Electron microscopy reveals profound nuclear changes culminating in nuclear dissolution.
- **Fates of necrotic cells.** Necrotic cells may persist for some time or may be digested by enzymes and disappear. Dead cells may be replaced by myelin figures, which are either phagocytosed by other cells or further degraded into fatty acids. These fatty acids bind calcium salts, which may result in the dead cells ultimately becoming **calcified.**

Patterns of Tissue Necrosis

There are several morphologically distinct patterns of tissue necrosis, which may provide clues about the underlying cause. Although the terms that describe these patterns do not reflect underlying mechanisms, such terms are in common use, and their implications are understood by both pathologists and clinicians. Most of these types of necrosis have distinct gross appearance; fibrinoid necrosis is detected only by histologic examination.

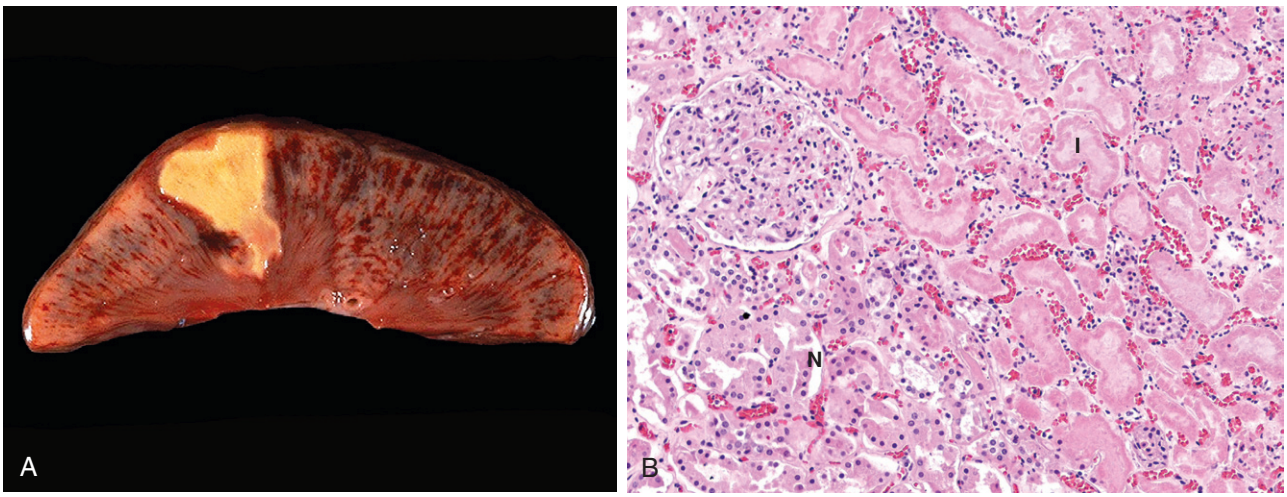


Figure 1–9 Coagulative necrosis. **A,** A wedge-shaped kidney infarct *(yellow)* with preservation of the outlines. **B,** Microscopic view of the edge of the infarct, with normal kidney *(N)* and necrotic cells in the infarct *(I)*. The necrotic cells show preserved outlines with loss of nuclei, and an inflammatory infiltrate is present (difficult to discern at this magnification).

MORPHOLOGY

- **Coagulative necrosis** is a form of necrosis in which the underlying tissue architecture is preserved for at least several days (Fig. 1–9). The affected tissues take on a firm texture. Presumably the injury denatures not only structural proteins but also enzymes, thereby blocking the proteolysis of the dead cells; as a result, eosinophilic, anucleate cells may persist for days or weeks. Leukocytes are recruited to the site of necrosis, and the dead cells are digested by the action of lysosomal enzymes of the leukocytes. The cellular debris is then removed by phagocytosis. Coagulative necrosis is characteristic of **infarcts** (areas of ischemic necrosis) in all of the solid organs except the brain.
- **Liquefactive necrosis** is seen in focal bacterial or, occasionally, fungal infections, because microbes stimulate the accumulation of inflammatory cells and the enzymes of leukocytes digest ("liquefy") the tissue. For obscure reasons, hypoxic death of cells within the central nervous system often evokes liquefactive necrosis (Fig. 1–10). Whatever the pathogenesis, the dead cells are completely digested, transforming the tissue into a liquid viscous mass. Eventually, the digested tissue is removed by phagocytes. If the process was initiated by acute inflammation, as in a bacterial infection, the material is frequently creamy yellow and is called pus (Chapter 2).
- Although **gangrenous necrosis** is not a distinctive pattern of cell death, the term is still commonly used in clinical practice. It usually refers to the condition of a limb, generally the lower leg, that has lost its blood supply and has undergone coagulative necrosis involving multiple tissue layers. When bacterial infection is superimposed, coagulative necrosis is modified by the liquefactive action of the bacteria and the attracted leukocytes (resulting in so-called **wet gangrene**).
- **Caseous necrosis** is encountered most often in foci of tuberculous infection. **Caseous** means "cheese-like," referring to the friable yellow-white appearance of the area of necrosis (Fig. 1–11). On microscopic examination, the necrotic focus appears as a collection of fragmented or lysed cells with an amorphous granular pink appearance in the usual H&E-stained tissue. Unlike with coagulative necrosis, the tissue architecture is completely obliterated and cellular outlines cannot be discerned. The area of caseous necrosis is often enclosed within a distinctive inflammatory border; this appearance is characteristic of a focus of inflammation known as a **granuloma** (Chapter 2).
- **Fat necrosis** refers to focal areas of fat destruction, typically resulting from release of activated pancreatic lipases into the substance of the pancreas and the peritoneal cavity. This occurs in the calamitous abdominal emergency known as acute pancreatitis (Chapter 16). In this disorder, pancreatic enzymes that have leaked out of acinar cells

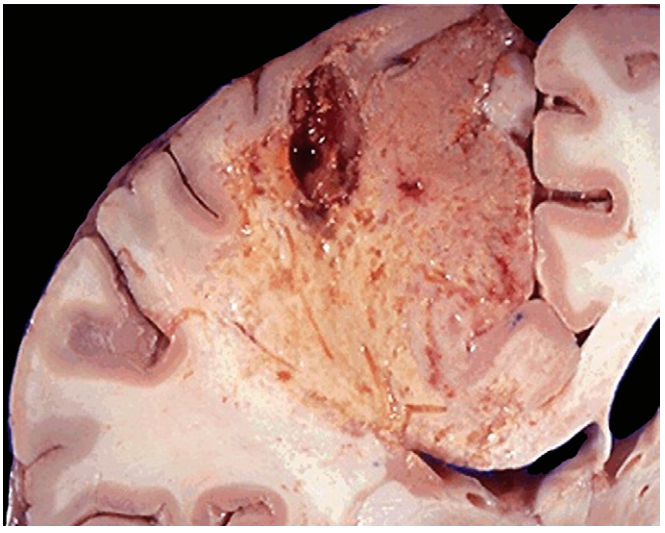


Figure 1–10 Liquefactive necrosis. An infarct in the brain showing dissolution of the tissue.

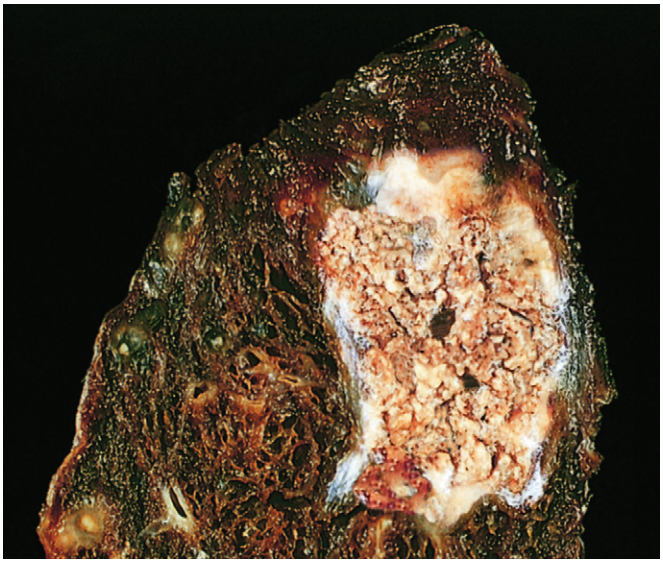


Figure 1–11 Caseous necrosis. Tuberculosis of the lung, with a large area of caseous necrosis containing yellow-white (cheesy) debris.

and ducts liquefy the membranes of fat cells in the peritoneum, and lipases split the triglyceride esters contained within fat cells. The released fatty acids combine with calcium to produce grossly visible chalky white areas (fat saponification), which enable the surgeon and the pathologist to identify the lesions (Fig. 1–12). On histologic examination, the foci of necrosis contain shadowy outlines of necrotic fat cells with basophilic calcium deposits, surrounded by an inflammatory reaction.

- **Fibrinoid necrosis** is a special form of necrosis, visible by light microscopy, usually in immune reactions in which complexes of antigens and antibodies are deposited in the walls of arteries. The deposited immune complexes, together with fibrin that has leaked out of vessels, produce a bright pink and amorphous appearance on H&E preparations called **fibrinoid** (fibrin-like) by pathologists (Fig. 1–13). The immunologically mediated diseases (e.g., polyarteritis nodosa) in which this type of necrosis is seen are described in Chapter 4.

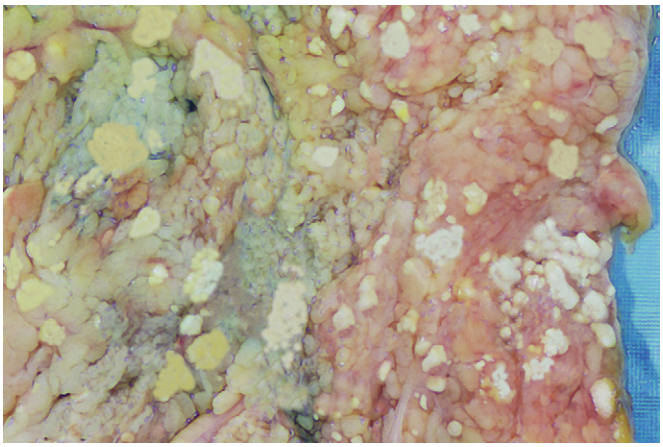


Figure 1–12 Fat necrosis in acute pancreatitis. The areas of white chalky deposits represent foci of fat necrosis with calcium soap formation (saponification) at sites of lipid breakdown in the mesentery.

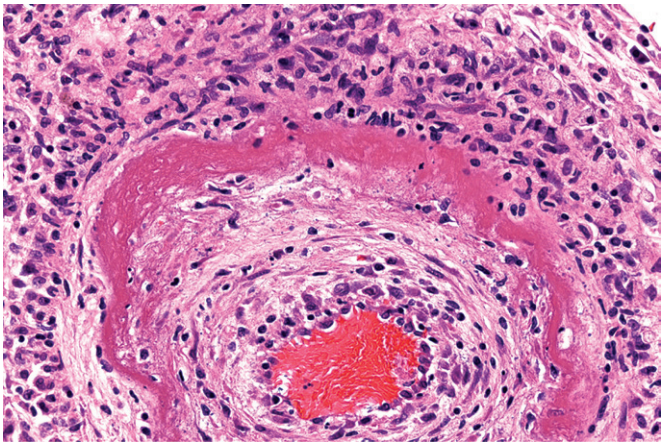


Figure 1–13 Fibrinoid necrosis in an artery in a patient with polyarteritis nodosa. The wall of the artery shows a circumferential bright pink area of necrosis with protein deposition and inflammation.

Leakage of intracellular proteins through the damaged cell membrane and ultimately into the circulation provides a means of detecting tissue-specific necrosis using blood or serum samples. Cardiac muscle, for example, contains a unique isoform of the enzyme creatine kinase and of the contractile protein troponin, whereas hepatic bile duct epithelium contains a temperature-resistant isoform of the enzyme alkaline phosphatase, and hepatocytes contain transaminases. Irreversible injury and cell death in these tissues result in increased serum levels of such proteins, and measurement of serum levels is used clinically to assess damage to these tissues.

SUMMARY

Morphologic Alterations in Injured Cells and Tissues

- *Reversible cell injury:* cell swelling, fatty change, plasma membrane blebbing and loss of microvilli, mitochondrial swelling, dilation of the ER, eosinophilia (due to decreased cytoplasmic RNA)
- *Necrosis:* increased eosinophilia; nuclear shrinkage, fragmentation, and dissolution; breakdown of plasma membrane and organellar membranes; abundant myelin figures; leakage and enzymatic digestion of cellular contents
- *Patterns of tissue necrosis:* Under different conditions, necrosis in tissues may assume specific patterns: coagulative, liquefactive, gangrenous, caseous, fat, and fibrinoid.

MECHANISMS OF CELL INJURY

Now that we have discussed the causes of cell injury and the morphologic changes in necrosis, we next consider in more detail the molecular basis of cell injury, and then illustrate the important principles with a few selected examples of common types of injury.

The biochemical mechanisms linking any given injury with the resulting cellular and tissue manifestations are complex, interconnected, and tightly interwoven with many intracellular metabolic pathways. Nevertheless,

several general principles are relevant to most forms of cell injury:

- *The cellular response to injurious stimuli depends on the type of injury, its duration, and its severity.* Thus, low doses of toxins or a brief duration of ischemia may lead to reversible cell injury, whereas larger toxin doses or longer ischemic intervals may result in irreversible injury and cell death.
- *The consequences of an injurious stimulus depend on the type, status, adaptability, and genetic makeup of the injured cell.* The same injury has vastly different outcomes depending on the cell type; thus, striated skeletal muscle in the leg accommodates complete ischemia for 2 to 3 hours without irreversible injury, whereas cardiac muscle dies after only 20 to 30 minutes. The nutritional (or hormonal) status can also be important; clearly, a glycogen-replete hepatocyte will tolerate ischemia much better than one that has just burned its last glucose molecule. Genetically determined diversity in metabolic pathways can contribute to differences in responses to injurious stimuli. For instance, when exposed to the same dose of a toxin, individuals who inherit variants in genes encoding cytochrome P-450 may catabolize the toxin at different rates, leading to different outcomes. Much effort is now directed toward understanding the role of genetic polymorphisms in responses to drugs and toxins. The study of such interactions is called pharmacogenomics. In fact, genetic variations influence susceptibility to many complex diseases as well as responsiveness to various therapeutic agents. Using the genetic makeup of the individual patient to guide therapy is one example of "personalized medicine."
- *Cell injury results from functional and biochemical abnormalities in one or more of several essential cellular components* (Fig. 1–14). The principal targets and biochemical mechanisms of cell injury are: (1) mitochondria and their ability to generate ATP and ROS under pathologic conditions; (2) disturbance in calcium homeostasis; (3) damage to cellular (plasma and lysosomal) membranes; and (4) damage to DNA and misfolding of proteins.
- *Multiple biochemical alterations may be triggered by any one injurious insult.* It is therefore difficult to assign any one mechanism to a particular insult or clinical situation in which cell injury is prominent. For this reason, therapies that target individual mechanisms of cell injury may not be effective.

With this background, we can briefly discuss the major biochemical mechanisms of cell injury.

Depletion of ATP

ATP, the energy store of cells, is produced mainly by oxidative phosphorylation of adenosine diphosphate (ADP) during reduction of oxygen in the electron transport system of mitochondria. In addition, the glycolytic pathway can generate ATP in the absence of oxygen using glucose derived either from the circulation or from the hydrolysis of intracellular glycogen. The major causes of ATP depletion are reduced supply of oxygen and nutrients, mitochondrial damage, and the actions of some toxins (e.g., cyanide). Tissues with a greater glycolytic capacity (e.g., the liver) are able to survive loss of oxygen and decreased oxidative phosphorylation better than are tissues with limited capacity for glycolysis (e.g., the brain). High-energy phosphate in the form of ATP is required for virtually all synthetic and degradative processes within the cell, including membrane transport, protein synthesis, lipogenesis, and the deacylation-reacylation reactions necessary for phospholipid turnover. It is estimated that in total, the cells of a healthy human burn 50 to 75 kg of ATP every day!

Significant depletion of ATP has widespread effects on many critical cellular systems (Fig. 1–15):

- The activity of *plasma membrane ATP-dependent sodium pumps* is reduced, resulting in intracellular accumulation of sodium and efflux of potassium. The net gain of solute is accompanied by iso-osmotic gain of water, causing *cell swelling* and dilation of the ER.
- There is a compensatory *increase in anaerobic glycolysis* in an attempt to maintain the cell's energy sources. As a consequence, intracellular glycogen stores are rapidly depleted, and lactic acid accumulates, leading to decreased intracellular pH and decreased activity of many cellular enzymes.
- *Failure of ATP-dependent Ca^{2+} pumps* leads to influx of Ca^{2+}, with damaging effects on numerous cellular components, described later.

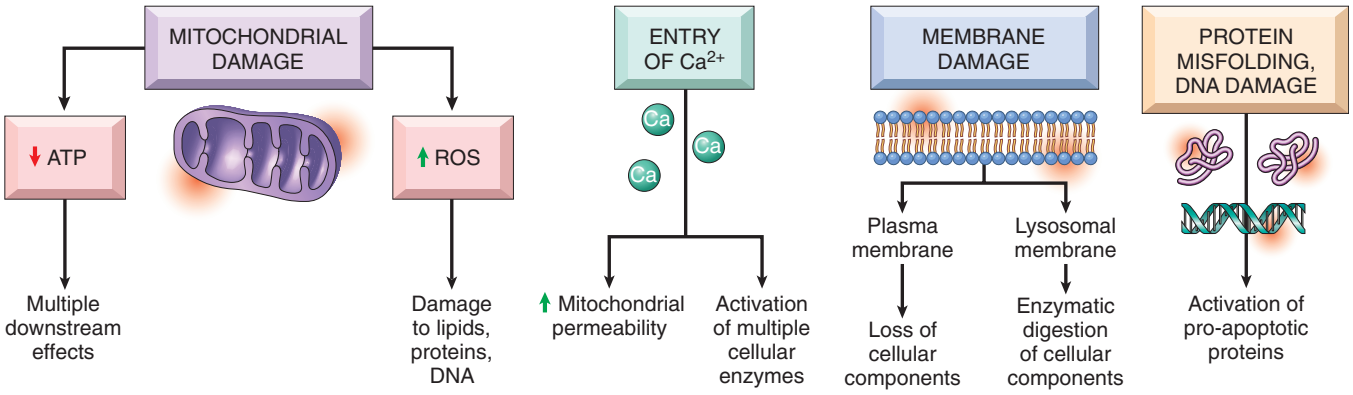


Figure 1–14 The principal biochemical mechanisms and sites of damage in cell injury. ATP, adenosine triphospate; ROS, reactive oxygen species.

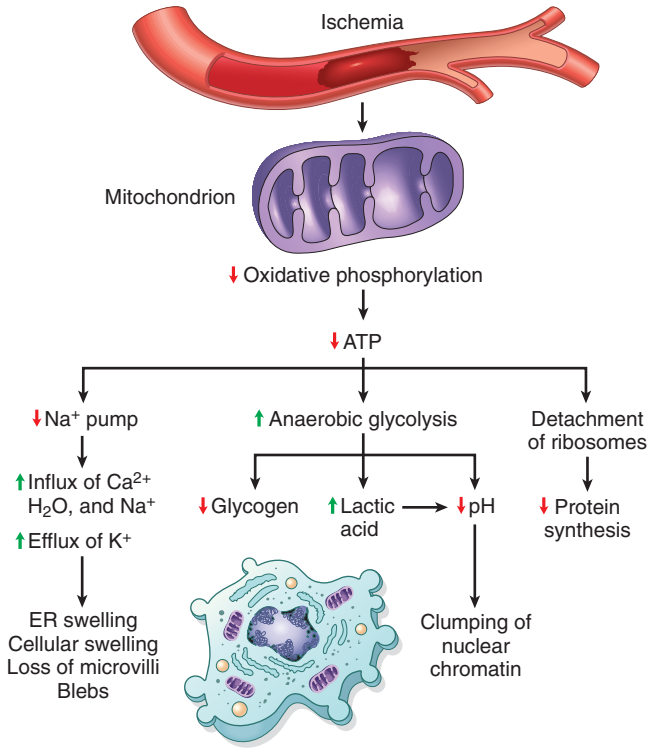


Figure 1–15 The functional and morphologic consequences of depletion of intracellular adenosine triphosphate (ATP). ER, endoplasmic reticulum.

- Prolonged or worsening depletion of ATP causes *structural disruption of the protein synthetic apparatus,* manifested as detachment of ribosomes from the rough ER (RER) and dissociation of polysomes into monosomes, with a consequent reduction in protein synthesis. Ultimately, there is irreversible damage to mitochondrial and lysosomal membranes, and the cell undergoes necrosis.

Mitochondrial Damage and Dysfunction

Mitochondria may be viewed as "mini-factories" that produce life-sustaining energy in the form of ATP. Not surprisingly, therefore, they are also critical players in cell injury and death (Fig. 1–16). Mitochondria are sensitive to many types of injurious stimuli, including hypoxia, chemical toxins, and radiation. Mitochondrial damage may result in several biochemical abnormalities:

- Failure of oxidative phosphorylation leads to progressive depletion of ATP, culminating in necrosis of the cell, as described earlier.
- Abnormal oxidative phosphorylation also leads to the formation of reactive oxygen species, which have many deleterious effects, described below.
- Damage to mitochondria is often associated with the formation of a high-conductance channel in the mitochondrial membrane, called the mitochondrial permeability transition pore. The opening of this channel leads to the loss of mitochondrial membrane potential and pH changes, further compromising oxidative phosphorylation.
- The mitochondria also contain several proteins that, when released into the cytoplasm, tell the cell there is internal injury and activate a pathway of apoptosis, discussed later.

Influx of Calcium

The importance of Ca^{2+} in cell injury was established by the experimental finding that depleting extracellular Ca^{2+} delays cell death after hypoxia and exposure to some toxins. Cytosolic free calcium is normally maintained by ATP-dependent calcium transporters at concentrations as much as 10,000 times lower than the concentration of extracellular calcium or of sequestered intracellular mitochondrial and ER calcium. Ischemia and certain toxins cause an increase in cytosolic calcium concentration, initially because of release of Ca^{2+} from the intracellular stores, and later resulting from increased influx across the plasma membrane. *Increased cytosolic Ca^{2+} activates a number of enzymes,* with potentially deleterious cellular effects (Fig. 1–17). These enzymes include phospholipases (which cause membrane damage), proteases (which break down both membrane and cytoskeletal proteins), endonucleases (which are responsible for DNA and chromatin fragmentation), and adenosine triphosphatases (ATPases) (thereby hastening ATP depletion). Increased intracellular Ca^{2+} levels may also induce apoptosis, by direct activation of caspases and by increasing mitochondrial permeability.

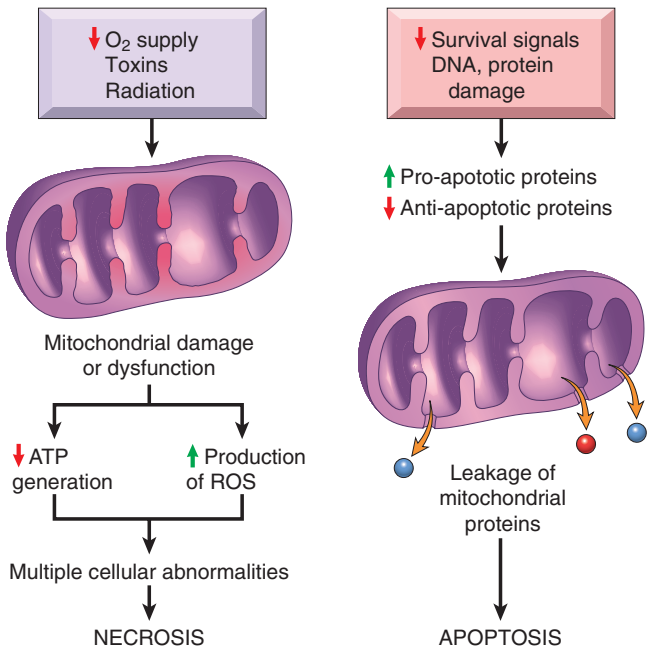


Figure 1–16 Role of mitochondria in cell injury and death. Mitochondria are affected by a variety of injurious stimuli and their abnormalities lead to necrosis or apoptosis. This pathway of apoptosis is described in more detail later. ATP, adenosine triphosphate; ROS, reactive oxygen species.

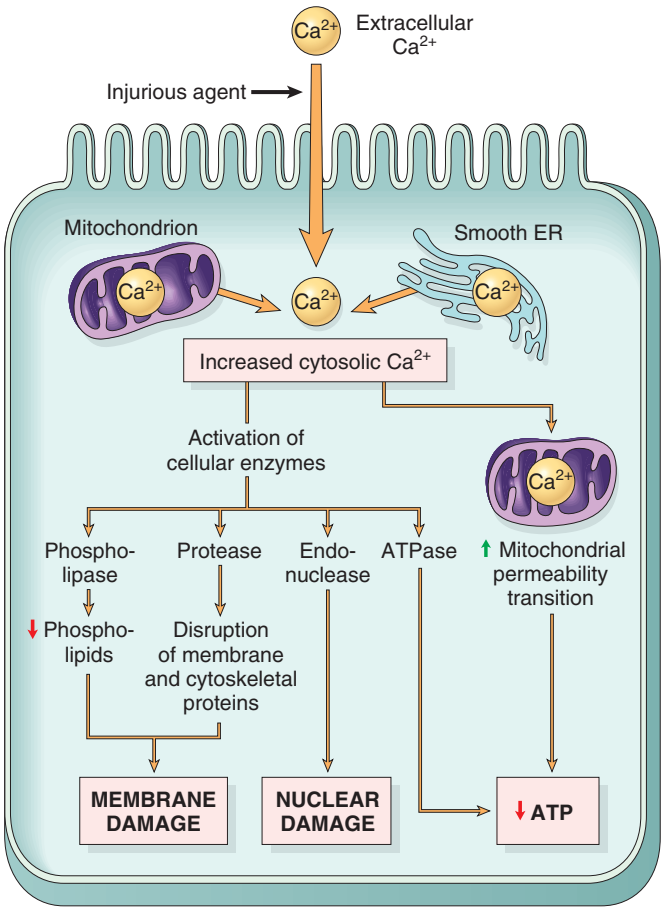


Figure 1–17 Sources and consequences of increased cytosolic calcium in cell injury. ATP, adenosine triphosphate; ATPase, adenosine triphosphatase.

Accumulation of Oxygen-Derived Free Radicals (Oxidative Stress)

Free radicals are chemical species with a single unpaired electron in an outer orbital. Such chemical states are extremely unstable, and free radicals readily react with inorganic and organic chemicals; when generated in cells, they avidly attack nucleic acids as well as a variety of cellular proteins and lipids. In addition, free radicals initiate reactions in which molecules that react with free radicals are themselves converted into other types of free radicals, thereby propagating the chain of damage.

Reactive oxygen species (ROS) are a type of oxygen-derived free radical whose role in cell injury is well established. Cell injury in many circumstances involves damage by free radicals; these situations include ischemia-reperfusion (discussed later on), chemical and radiation injury, toxicity from oxygen and other gases, cellular aging, microbial killing by phagocytic cells, and tissue injury caused by inflammatory cells.

There are different types of ROS, and they are produced by two major pathways (Fig. 1–18).

- *ROS are produced normally in small amounts in all cells during the reduction-oxidation (redox) reactions* that occur during mitochondrial respiration and energy generation. In this process, molecular oxygen is sequentially reduced in mitochondria by the addition of four electrons to generate water. This reaction is imperfect, however, and small amounts of highly reactive but short-lived toxic intermediates are generated when oxygen is only partially reduced. These intermediates include superoxide ($O_2^{\cdot-}$), which is converted to hydrogen peroxide (H_2O_2) spontaneously and by the action of the enzyme superoxide dismutase. H_2O_2 is more stable than $O_2^{\cdot-}$ and can cross biologic membranes. In the presence of metals, such as Fe^{2+}, H_2O_2 is converted to the highly reactive hydroxyl radical $^{\cdot}OH$ by the Fenton reaction.

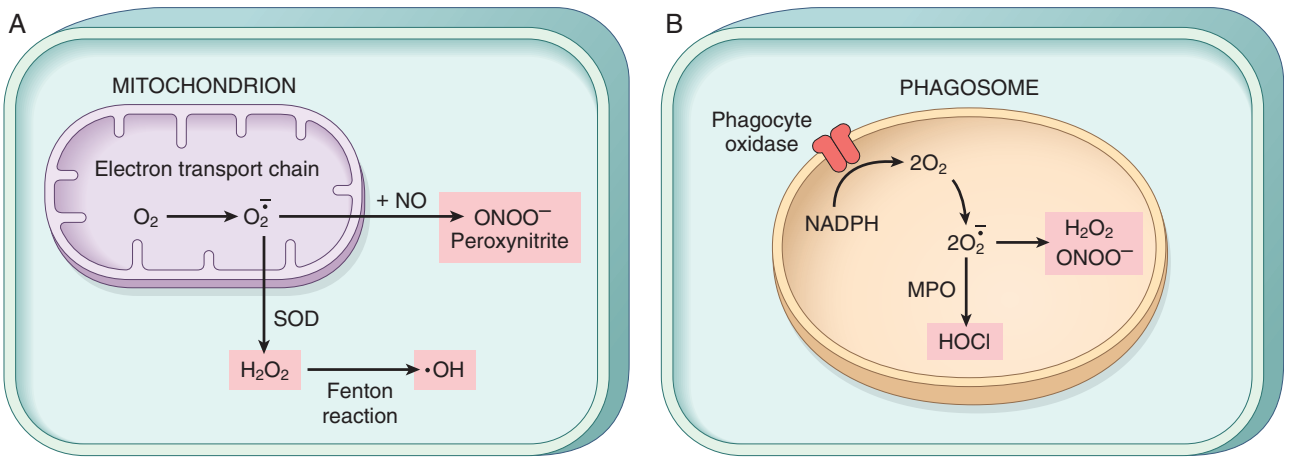


Figure 1–18 Pathways of production of reactive oxygen species. **A,** In all cells, superoxide ($O_2^{\cdot-}$) is generated during mitochondrial respiration by the electron transport chain and may be converted to H_2O_2 and the hydroxyl ($^{\cdot}OH$) free radical or to peroxynitrite ($ONOO^-$). **B,** In leukocytes (mainly neutrophils and macrophages), the phagocyte oxidase enzyme in the phagosome membrane generates superoxide, which can be converted to other free radicals. Myeloperoxidase (MPO) in phagosomes also generates hypochlorite from reactive oxygen species (ROS). NO, nitric oxide; SOD, superoxide dismutase.

- *ROS are produced in phagocytic leukocytes, mainly neutrophils and macrophages,* as a weapon for destroying ingested microbes and other substances during inflammation and host defense (Chapter 2). The ROS are generated in the phagosomes and phagolysosomes of leukocytes by a process that is similar to mitochondrial respiration and is called the *respiratory burst* (or oxidative burst). In this process, a phagosome membrane enzyme catalyzes the generation of superoxide, which is converted to H_2O_2. H_2O_2 is in turn converted to a highly reactive compound hypochlorite (the major component of household bleach) by the enzyme myeloperoxidase, which is present in leukocytes. The role of ROS in inflammation is described in Chapter 2.
- *Nitric oxide (NO)* is another reactive free radical produced in leukocytes and other cells. It can react with $O_2^{\bullet-}$ to form a highly reactive compound, peroxynitrite, which also participates in cell injury.

The damage caused by free radicals is determined by their rates of production and removal (Fig. 1–19). When the production of ROS increases or the scavenging systems are ineffective, the result is an excess of these free radicals, leading to a condition called *oxidative stress.*

The generation of free radicals is increased under several circumstances:

- The absorption of radiant energy (e.g., ultraviolet light, x-rays). Ionizing radiation can hydrolyze water into hydroxyl ($^{\bullet}OH$) and hydrogen ($H^{\bullet}$) free radicals.
- The enzymatic metabolism of exogenous chemicals (e.g., carbon tetrachloride—see later)
- Inflammation, in which free radicals are produced by leukocytes (Chapter 2)

Cells have developed many *mechanisms to remove free radicals* and thereby minimize injury. Free radicals are inherently unstable and decay spontaneously. There are also nonenzymatic and enzymatic systems that contribute to inactivation of free radicals (Fig. 1–19).

- The rate of decay of superoxide is significantly increased by the action of superoxide dismutases (SODs) found in many cell types.
- Glutathione (GSH) peroxidases are a family of enzymes whose major function is to protect cells from oxidative damage. The most abundant member of this family, glutathione peroxidase 1, is found in the cytoplasm of all cells. It catalyzes the breakdown of H_2O_2 by the reaction 2 GSH (glutathione) + H_2O_2 → GS-SG + 2 H_2O. The intracellular ratio of oxidized glutathione (GSSG) to reduced glutathione (GSH) is a reflection of this enzyme's activity and thus of the cell's ability to catabolize free radicals.
- Catalase, present in peroxisomes, catalyzes the decomposition of hydrogen peroxide ($2H_2O_2 \rightarrow O_2 + 2H_2O$). It is one of the most active enzymes known, capable of degrading millions of molecules of H_2O_2 per second.
- Endogenous or exogenous antioxidants (e.g., vitamins E, A, and C and β-carotene) may either block the formation of free radicals or scavenge them once they have formed.

Reactive oxygen species cause cell injury by three main reactions (Fig. 1–19):

- *Lipid peroxidation of membranes.* Double bonds in membrane polyunsaturated lipids are vulnerable to attack by oxygen-derived free radicals. The lipid–radical interactions yield peroxides, which are themselves unstable and reactive, and an autocatalytic chain reaction ensues.
- *Cross-linking and other changes in proteins.* Free radicals promote sulfhydryl-mediated protein cross-linking, resulting in enhanced degradation or loss of enzymatic activity. Free radical reactions may also directly cause polypeptide fragmentation.
- *DNA damage.* Free radical reactions with thymine in nuclear and mitochondrial DNA produce single-strand breaks. Such DNA damage has been implicated in cell death, aging, and malignant transformation of cells.

In addition to the role of ROS in cell injury and killing of microbes, low concentrations of ROS are involved in numerous signaling pathways in cells and thus in many physiologic reactions. Therefore, these molecules are produced normally but, to avoid their harmful effects, their intracellular concentrations are tightly regulated in healthy cells.

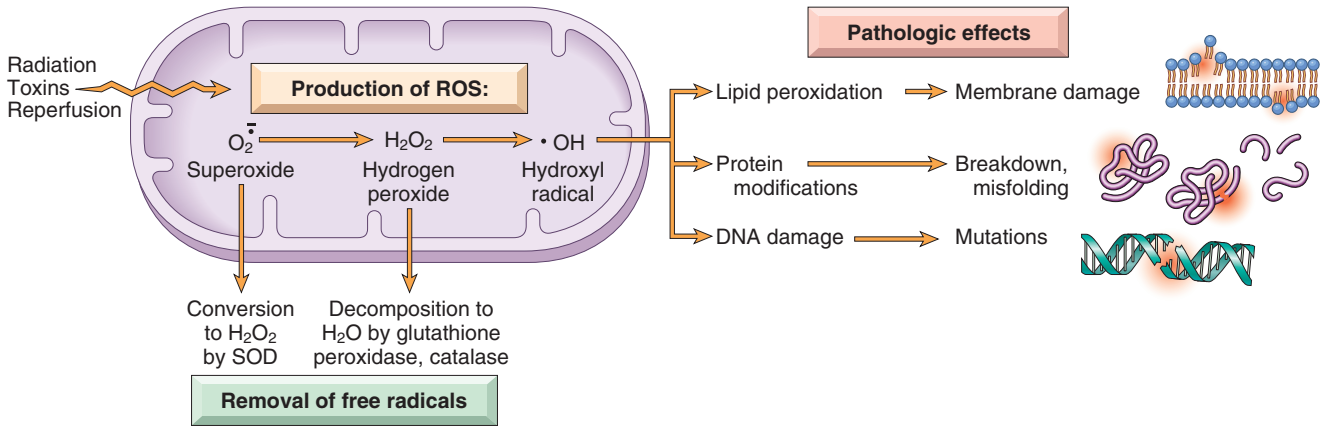


Figure 1–19 The generation, removal, and role of reactive oxygen species (ROS) in cell injury. The production of ROS is increased by many injurious stimuli. These free radicals are removed by spontaneous decay and by specialized enzymatic systems. Excessive production or inadequate removal leads to accumulation of free radicals in cells, which may damage lipids (by peroxidation), proteins, and deoxyribonucleic acid (DNA), resulting in cell injury.

Defects in Membrane Permeability

Increased membrane permeability leading ultimately to overt membrane damage is a consistent feature of most forms of cell injury that culminate in necrosis. The plasma membrane can be damaged by ischemia, various microbial toxins, lytic complement components, and a variety of physical and chemical agents. Several biochemical mechanisms may contribute to membrane damage (Fig. 1–20):

- *Decreased phospholipid synthesis.* The production of phospholipids in cells may be reduced whenever there is a fall in ATP levels, leading to decreased energy-dependent enzymatic activities. The reduced phospholipid synthesis may affect all cellular membranes, including the membranes of mitochondria, thus exacerbating the loss of ATP.
- *Increased phospholipid breakdown.* Severe cell injury is associated with increased degradation of membrane phospholipids, probably owing to activation of endogenous phospholipases by increased levels of cytosolic Ca^{2+}.
- *ROS.* Oxygen free radicals cause injury to cell membranes by lipid peroxidation, discussed earlier.
- *Cytoskeletal abnormalities.* Cytoskeletal filaments act as anchors connecting the plasma membrane to the cell interior, and serve many functions in maintaining normal cellular architecture, motility, and signaling. Activation of proteases by increased cytosolic Ca^{2+} may cause damage to elements of the cytoskeleton, leading to membrane damage.
- *Lipid breakdown products.* These include unesterified free fatty acids, acyl carnitine, and lysophospholipids, all of which accumulate in injured cells as a result of phospholipid degradation. These catabolic products have a detergent effect on membranes. They may also either insert into the lipid bilayer of the membrane or exchange with membrane phospholipids, causing changes in permeability and electrophysiologic alterations.

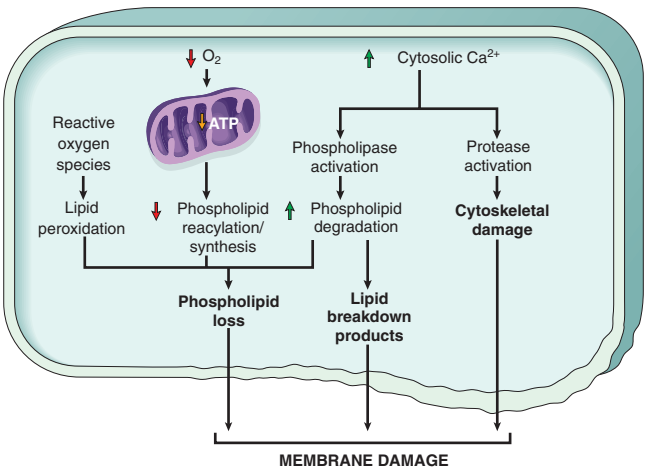


Figure 1–20 Mechanisms of membrane damage in cell injury. Decreased O_2 and increased cytosolic Ca^{2+} typically are seen in ischemia but may accompany other forms of cell injury. Reactive oxygen species, which often are produced on reperfusion of ischemic tissues, also cause membrane damage (*not shown*).

The most important sites of membrane damage during cell injury are the mitochondrial membrane, the plasma membrane, and membranes of lysosomes.

- *Mitochondrial membrane damage.* As discussed earlier, damage to mitochondrial membranes results in decreased production of ATP, with many deleterious effects culminating in necrosis.
- *Plasma membrane damage.* Plasma membrane damage leads to loss of osmotic balance and influx of fluids and ions, as well as loss of cellular contents. The cells may also leak metabolites that are vital for the reconstitution of ATP, thus further depleting energy stores.
- *Injury to lysosomal membranes* results in leakage of their enzymes into the cytoplasm and activation of the acid hydrolases in the acidic intracellular pH of the injured (e.g., ischemic) cell. Lysosomes contain ribonucleases (RNases), DNases, proteases, glucosidases, and other enzymes. Activation of these enzymes leads to enzymatic digestion of cell components, and the cells die by necrosis.

Damage to DNA and Proteins

Cells have mechanisms that repair damage to DNA, but if this damage is too severe to be corrected (e.g., after radiation injury or oxidative stress), the cell initiates its suicide program and dies by apoptosis. A similar reaction is triggered by the accumulation of improperly folded proteins, which may result from inherited mutations or external triggers such as free radicals. Since these mechanisms of cell injury typically cause apoptosis, they are discussed later in the chapter.

SUMMARY

Mechanisms of Cell Injury

- *ATP depletion:* failure of energy-dependent functions → reversible injury → necrosis
- *Mitochondrial damage:* ATP depletion → failure of energy-dependent cellular functions → ultimately, necrosis; under some conditions, leakage of mitochondrial proteins that cause apoptosis
- *Influx of calcium:* activation of enzymes that damage cellular components and may also trigger apoptosis
- *Accumulation of reactive oxygen species:* covalent modification of cellular proteins, lipids, nucleic acids
- *Increased permeability of cellular membranes:* may affect plasma membrane, lysosomal membranes, mitochondrial membranes; typically culminates in necrosis
- *Accumulation of damaged DNA and misfolded proteins:* triggers apoptosis

CLINICOPATHOLOGIC CORRELATIONS: EXAMPLES OF CELL INJURY AND NECROSIS

To illustrate the evolution and biochemical mechanisms of cell injury, we conclude this section by discussing some commonly encountered examples of reversible cell injury and necrosis.

Ischemic and Hypoxic Injury

Ischemia, or diminished blood flow to a tissue, is a common cause of acute cell injury underlying human disease. In contrast with hypoxia, in which energy generation by anaerobic glycolysis can continue (albeit less efficiently than by oxidative pathways), ischemia, because of reduced blood supply, also compromises the delivery of substrates for glycolysis. Consequently, anaerobic energy generation also ceases in ischemic tissues after potential substrates are exhausted or when glycolysis is inhibited by the accumulation of metabolites that would normally be removed by blood flow. Therefore, *ischemia injures tissues faster and usually more severely than does hypoxia.* The major cellular abnormalities in oxygen-deprived cells are decreased ATP generation, mitochondrial damage, and accumulation of ROS, with its downstream consequences.

The most important biochemical abnormality in hypoxic cells that leads to cell injury is reduced intracellular generation of ATP, as a consequence of reduced supply of oxygen. As described above, loss of ATP leads to the failure of many energy-dependent cellular systems, including (1) ion pumps (leading to cell swelling, and influx of Ca^{2+}, with its deleterious consequences); (2) depletion of glycogen stores and accumulation of lactic acid, thus lowering the intracellular pH; and (3) reduction in protein synthesis.

The functional consequences may be severe at this stage. For instance, heart muscle ceases to contract within 60 seconds of coronary occlusion. If hypoxia continues, worsening ATP depletion causes further deterioration, with loss of microvilli and the formation of "blebs" (Fig. 1–6). At this time, the entire cell and its organelles (mitochondria, ER) are markedly swollen, with increased concentrations of water, sodium, and chloride and a decreased concentration of potassium. *If oxygen is restored, all of these disturbances are reversible,* and in the case of myocardium, contractility returns.

If ischemia persists, irreversible injury and necrosis ensue. Irreversible injury is associated with severe swelling of mitochondria, extensive damage to plasma membranes, and swelling of lysosomes. ROS accumulate in cells, and massive influx of calcium may occur. Death is mainly by necrosis, but apoptosis also contributes; the apoptotic pathway is activated by release of pro-apoptotic molecules from mitochondria. The cell's components are progressively degraded, and there is widespread leakage of cellular enzymes into the extracellular space. Finally, the dead cells may become replaced by large masses composed of phospholipids in the form of myelin figures. These are then either phagocytosed by leukocytes or degraded further into fatty acids that may become calcified.

Ischemia-Reperfusion Injury

If cells are reversibly injured, the restoration of blood flow can result in cell recovery. However, under certain circumstances, *the restoration of blood flow to ischemic but viable tissues results, paradoxically, in the death of cells that are not otherwise irreversibly injured.* This so-called *ischemia-reperfusion injury* is a clinically important process that may contribute significantly to tissue damage in myocardial and cerebral ischemia.

Several mechanisms may account for the exacerbation of cell injury resulting from reperfusion into ischemic tissues:

- New damage may be initiated during reoxygenation by increased generation of *ROS* from parenchymal and endothelial cells and from infiltrating leukocytes. When the supply of oxygen is increased, there may be a corresponding increase in the production of ROS, especially because mitochondrial damage leads to incomplete reduction of oxygen, and because of the action of oxidases in leukocytes, endothelial cells, or parenchymal cells. Cellular antioxidant defense mechanisms may also be compromised by ischemia, favoring the accumulation of free radicals.
- The *inflammation* that is induced by ischemic injury may increase with reperfusion because of increased influx of leukocytes and plasma proteins. The products of activated leukocytes may cause additional tissue injury (Chapter 2). Activation of the *complement system* may also contribute to ischemia-reperfusion injury. Complement proteins may bind in the injured tissues, or to antibodies that are deposited in the tissues, and subsequent complement activation generates by-products that exacerbate the cell injury and inflammation.

Chemical (Toxic) Injury

Chemicals induce cell injury by one of two general mechanisms:

- *Some chemicals act directly by combining with a critical molecular component or cellular organelle.* For example, in mercuric chloride poisoning (as may occur from ingestion of contaminated seafood) (Chapter 7), mercury binds to the sulfhydryl groups of various cell membrane proteins, causing inhibition of ATP-dependent transport and increased membrane permeability. Many antineoplastic chemotherapeutic agents also induce cell damage by direct cytotoxic effects. In such instances, the *greatest damage is sustained by the cells that use, absorb, excrete, or concentrate the compounds.*
- *Many other chemicals are not intrinsically biologically active but must be first converted to reactive toxic metabolites, which then act on target cells.* This modification is usually accomplished by the cytochrome P-450 in the smooth ER of the liver and other organs. Although the metabolites might cause membrane damage and cell injury by direct covalent binding to protein and lipids, the most important mechanism of cell injury involves the formation of free radicals. *Carbon tetrachloride* (CCl_4)—once widely used in the dry cleaning industry but now banned—and the analgesic *acetaminophen* belong in this category. The effect of CCl_4 is still instructive as an example of chemical injury. CCl_4 is converted to the toxic free radical $CCl_3^{\bullet}$, principally in the liver, and this free radical is the cause of cell injury, mainly by membrane phospholipid peroxidation. In less than 30 minutes after exposure to CCl_4, there is breakdown of ER membranes with a decline in hepatic protein synthesis of enzymes and plasma proteins; within 2 hours, swelling of the smooth ER and dissociation of ribosomes from the smooth ER have occurred. There is reduced lipid export from the hepatocytes, as a result of their inability to synthesize

apoprotein to form complexes with triglycerides and thereby facilitate lipoprotein secretion; the result is the "fatty liver" of CCl_4 poisoning. Mitochondrial injury follows, and subsequently diminished ATP stores result in defective ion transport and progressive cell swelling; the plasma membranes are further damaged by fatty aldehydes produced by lipid peroxidation in the ER. The end result can be calcium influx and eventually cell death.

APOPTOSIS

Apoptosis is a pathway of cell death in which cells activate enzymes that degrade the cells' own nuclear DNA and nuclear and cytoplasmic proteins. Fragments of the apoptotic cells then break off, giving the appearance that is responsible for the name (*apoptosis,* "falling off"). The plasma membrane of the apoptotic cell remains intact, but the membrane is altered in such a way that the cell and its fragments become avid targets for phagocytes. The dead cell and its fragments are rapidly cleared before cellular contents have leaked out, so *apoptotic cell death does not elicit an inflammatory reaction in the host.* Apoptosis differs in this respect from necrosis, which is characterized by loss of membrane integrity, enzymatic digestion of cells, leakage of cellular contents, and frequently a host reaction (Fig. 1–6 and Table 1–1). However, apoptosis and necrosis sometimes coexist, and apoptosis induced by some pathologic stimuli may progress to necrosis.

Causes of Apoptosis

Apoptosis occurs in many normal situations and serves to eliminate potentially harmful cells and cells that have outlived their usefulness. It also occurs as a pathologic event when cells are damaged beyond repair, especially when the damage affects the cell's DNA or proteins; in these situations, the irreparably damaged cell is eliminated.

Apoptosis in Physiologic Situations

Death by apoptosis is a normal phenomenon that serves to eliminate cells that are no longer needed and to maintain a constant number of cells of various types in tissues. It is important in the following physiologic situations:

- *The programmed destruction of cells during embryogenesis.* Normal development is associated with the death of some cells and the appearance of new cells and tissues. The term *programmed cell death* was originally coined to denote this death of specific cell types at defined times during the development of an organism. Apoptosis is a generic term for this pattern of cell death, regardless of the context, but it is often used interchangeably with programmed cell death.
- *Involution of hormone-dependent tissues upon hormone deprivation,* such as endometrial cell breakdown during the menstrual cycle, and regression of the lactating breast after weaning
- *Cell loss in proliferating cell populations,* such as intestinal crypt epithelia, in order to maintain a constant number
- *Elimination of cells that have served their useful purpose,* such as neutrophils in an acute inflammatory response and lymphocytes at the end of an immune response. In these situations, cells undergo apoptosis because they are deprived of necessary survival signals, such as growth factors.
- *Elimination of potentially harmful self-reactive lymphocytes,* either before or after they have completed their maturation, in order to prevent reactions against the body's own tissues (Chapter 4)
- *Cell death induced by cytotoxic T lymphocytes,* a defense mechanism against viruses and tumors that serves to kill virus-infected and neoplastic cells (Chapter 4)

Apoptosis in Pathologic Conditions

Apoptosis eliminates cells that are genetically altered or injured beyond repair and does so without eliciting a severe host reaction, thereby keeping the extent of tissue damage to a minimum. Death by apoptosis is responsible for loss of cells in a variety of pathologic states:

- *DNA damage.* Radiation, cytotoxic anticancer drugs, extremes of temperature, and even hypoxia can damage DNA, either directly or through production of free radicals. If repair mechanisms cannot cope with the injury, the cell triggers intrinsic mechanisms that induce apoptosis. In these situations, elimination of the cell may be a better alternative than risking mutations in the damaged DNA, which may progress to malignant transformation. These injurious stimuli cause apoptosis if the insult is mild, but larger doses of the same stimuli result in necrotic cell death. Inducing apoptosis of cancer cells is a desired effect of chemotherapeutic agents, many of which work by damaging DNA.
- *Accumulation of misfolded proteins.* Improperly folded proteins may arise because of mutations in the genes encoding these proteins or because of extrinsic factors, such as damage caused by free radicals. Excessive accumulation of these proteins in the ER leads to a condition called *ER stress,* which culminates in apoptotic death of cells.
- *Cell injury in certain infections,* particularly viral infections, in which loss of infected cells is largely due to apoptotic death that may be induced by the virus (as in adenovirus and human immunodeficiency virus infections) or by the host immune response (as in viral hepatitis).
- *Pathologic atrophy in parenchymal organs after duct obstruction,* such as occurs in the pancreas, parotid gland, and kidney

MORPHOLOGY

In H&E-stained tissue sections, the nuclei of apoptotic cells show various stages of chromatin condensation and aggregation and, ultimately, karyorrhexis (Fig. 1–21); at the molecular level this is reflected in fragmentation of DNA into nucleosome-sized pieces. The cells rapidly shrink, form cytoplasmic buds, and fragment into **apoptotic bodies** composed of membrane-bound vesicles of cytosol and organelles (Fig. 1–6). Because these fragments are quickly extruded and phagocytosed without eliciting an inflammatory response, even substantial apoptosis may be histologically undetectable.

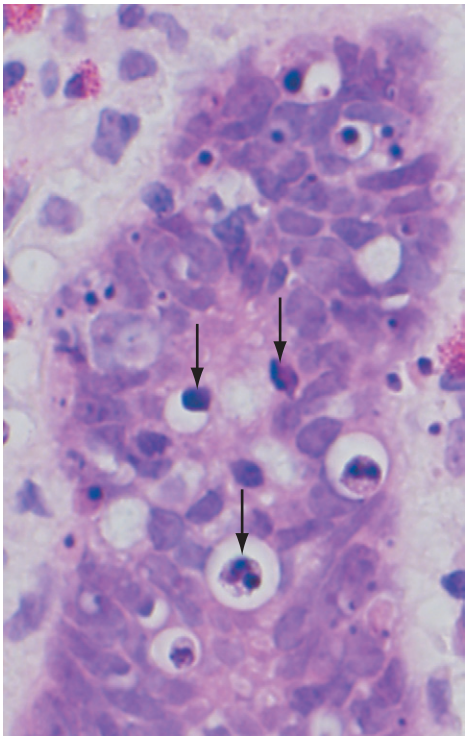


Figure 1–21 Morphologic appearance of apoptotic cells. Apoptotic cells (some indicated by *arrows*) in a normal crypt in the colonic epithelium are shown. (The preparative regimen for colonoscopy frequently induces apoptosis in epithelial cells, which explains the abundance of dead cells in this normal tissue.) Note the fragmented nuclei with condensed chromatin and the shrunken cell bodies, some with pieces falling off.
(Courtesy of Dr. Sanjay Kakar, Department of Pathology, University of California San Francisco, San Francisco, Calif.)

Mechanisms of Apoptosis

Apoptosis results from the activation of enzymes called caspases (so named because they are *c*ysteine proteases that cleave proteins after *asp*artic residues). The activation of caspases depends on a finely tuned balance between production of pro- and anti-apoptotic proteins. Two distinct pathways converge on caspase activation: the *mitochondrial pathway* and the *death receptor pathway* (Fig. 1–22). Although these pathways can intersect, they are generally induced under different conditions, involve different molecules, and serve distinct roles in physiology and disease.

The Mitochondrial (Intrinsic) Pathway of Apoptosis

Mitochondria contain several proteins that are capable of inducing apoptosis; these proteins include cytochrome c and other proteins that neutralize endogenous inhibitors of apoptosis. The choice between cell survival and death is determined by the permeability of mitochondria, which is controlled by a family of more than 20 proteins, the prototype of which is Bcl-2 (Fig. 1–23). When cells are deprived of growth factors and other survival signals, or are exposed to agents that damage DNA, or accumulate unacceptable amounts of misfolded proteins, a number of sensors are activated. These sensors are members of the Bcl-2 family called "BH3 proteins" (because they contain only the third of multiple conserved domains of the Bcl-2 family). They in turn activate two pro-apoptotic members of the family called Bax and Bak, which dimerize, insert into the mitochondrial membrane, and form channels through which cytochrome c and other mitochondrial proteins escape into the cytosol. These sensors also inhibit the anti-apoptotic molecules Bcl-2 and Bcl-x_L (see further on), enhancing the leakage of mitochondrial proteins. Cytochrome c, together with some cofactors, activates caspase-9. Other proteins that leak out of mitochondria block the activities of caspase antagonists that function as physiologic inhibitors of apoptosis. The net result is the activation of the caspase cascade, ultimately leading to nuclear fragmentation. Conversely, if cells are exposed to growth factors and other survival signals, they synthesize anti-apoptotic members of the Bcl-2 family, the two main ones of which are Bcl-2 itself and Bcl-x_L. These proteins antagonize Bax and Bak, and thus limit the escape of the mitochondrial pro-apoptotic proteins. Cells deprived of growth factors not only activate the pro-apoptotic Bax and Bak but also show reduced levels of Bcl-2 and Bcl-x_L, thus further tilting the balance toward death. The mitochondrial pathway seems to be the pathway that is responsible for apoptosis in most situations, as we discuss later.

The Death Receptor (Extrinsic) Pathway of Apoptosis

Many cells express surface molecules, called death receptors, that trigger apoptosis. Most of these are members of the tumor necrosis factor (TNF) receptor family, which contain in their cytoplasmic regions a conserved "death domain," so named because it mediates interaction with other proteins involved in cell death. The prototypic death receptors are the type I TNF receptor and Fas (CD95). Fas ligand (FasL) is a membrane protein expressed mainly on activated T lymphocytes. When these T cells recognize Fas-expressing targets, Fas molecules are cross-linked by FasL and bind adaptor proteins via the death domain. These in turn recruit and activate caspase-8. In many cell types caspase-8 may cleave and activate a pro-apoptotic member of the Bcl-2 family called Bid, thus feeding into the mitochondrial pathway. The combined activation of both pathways delivers a lethal blow to the cell. Cellular proteins, notably a caspase antagonist called FLIP, block activation of caspases downstream of death receptors. Interestingly, some viruses produce homologues of FLIP, and it is suggested that this is a mechanism that viruses use to keep infected cells alive. The death receptor pathway is involved in elimination of self-reactive lymphocytes and in killing of target cells by some cytotoxic T lymphocytes.

Activation and Function of Caspases

The mitochondrial and death receptor pathways lead to the activation of the *initiator caspases,* caspase-9 and -8, respectively. Active forms of these enzymes are produced, and these cleave and thereby activate another series of caspases that are called the *executioner caspases.* These activated caspases cleave numerous targets, culminating in activation of nucleases that degrade DNA and nucleoproteins. Caspases also degrade components of the nuclear matrix and cytoskeleton, leading to fragmentation of cells.

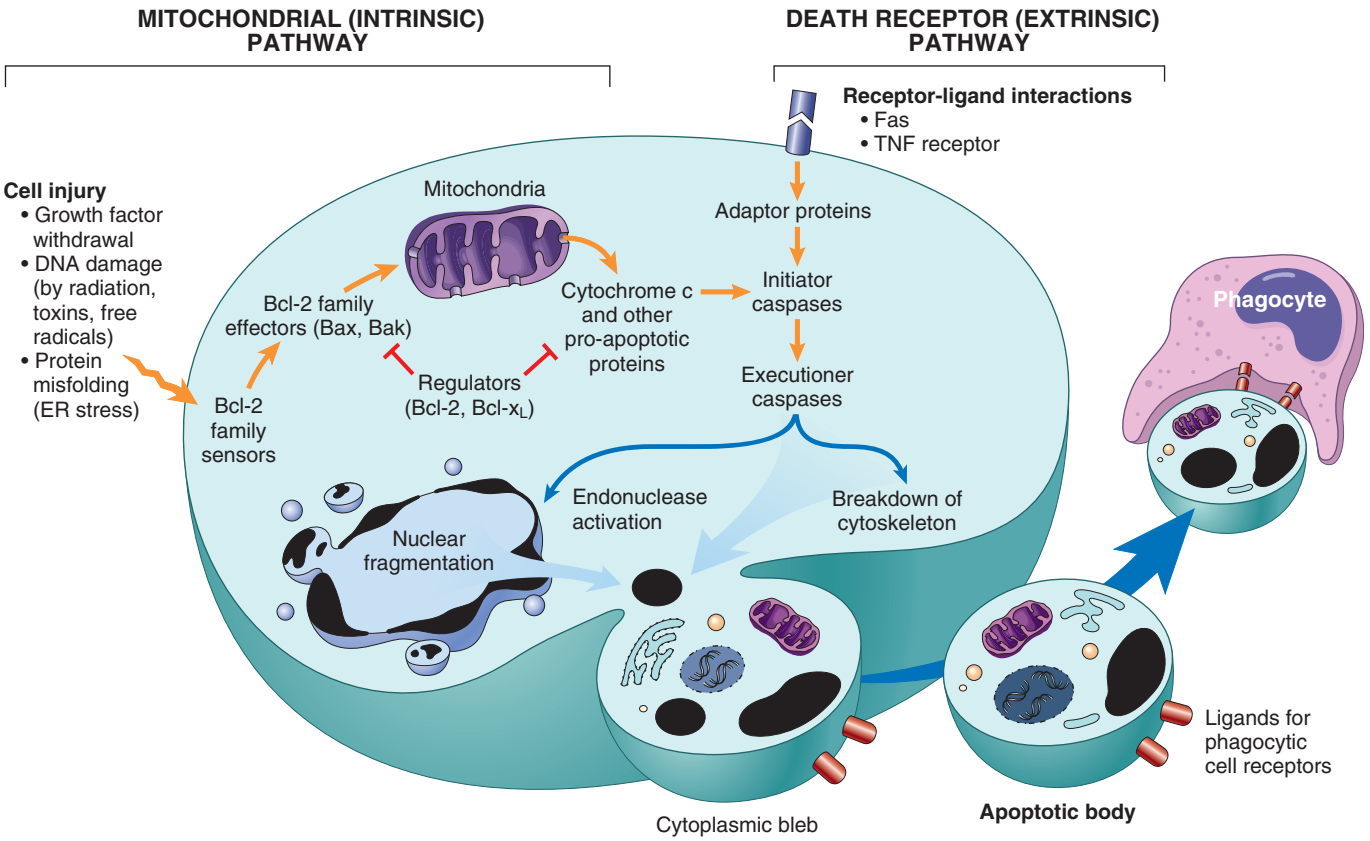


Figure 1–22 Mechanisms of apoptosis. The two pathways of apoptosis differ in their induction and regulation, and both culminate in the activation of caspases. In the mitochondrial pathway, proteins of the Bcl-2 family, which regulate mitochondrial permeability, become imbalanced and leakage of various substances from mitochondria leads to caspase activation. In the death receptor pathway, signals from plasma membrane receptors lead to the assembly of adaptor proteins into a "death-inducing signaling complex," which activates caspases, and the end result is the same.

Clearance of Apoptotic Cells

Apoptotic cells entice phagocytes by producing "eat-me" signals. In normal cells phosphatidylserine is present on the inner leaflet of the plasma membrane, but in apoptotic cells this phospholipid "flips" to the outer leaflet, where it is recognized by tissue macrophages and leads to phagocytosis of the apoptotic cells. Cells that are dying by apoptosis also secrete soluble factors that recruit phagocytes. This facilitates prompt clearance of the dead cells before they undergo secondary membrane damage and release their cellular contents (which can induce inflammation). Some apoptotic bodies express adhesive glycoproteins that are recognized by phagocytes, and macrophages themselves may produce proteins that bind to apoptotic cells (but not to live cells) and target the dead cells for engulfment. Numerous macrophage receptors have been shown to be involved in the binding and engulfment of apoptotic cells. This process of phagocytosis of apoptotic cells is so efficient that dead cells disappear without leaving a trace, and inflammation is virtually absent.

Although we have emphasized the distinctions between necrosis and apoptosis, these two forms of cell death may coexist and be related mechanistically. For instance, DNA damage (seen in apoptosis) activates an enzyme called poly-ADP(ribose) polymerase, which depletes cellular supplies of nicotinamide adenine dinucleotide, leading to a fall in ATP levels and ultimately necrosis. In fact, even in common situations such as ischemia, it has been suggested that early cell death can be partly attributed to apoptosis, with necrosis supervening later as ischemia worsens.

Examples of Apoptosis

Cell death in many situations is caused by apoptosis. The examples listed next illustrate the role of the two pathways of apoptosis in normal physiology and in disease.

Growth Factor Deprivation

Hormone-sensitive cells deprived of the relevant hormone, lymphocytes that are not stimulated by antigens and cytokines, and neurons deprived of nerve growth factor die by apoptosis. In all these situations, apoptosis is triggered by the mitochondrial pathway and is attributable to activation of pro-apoptotic members of the Bcl-2 family and decreased synthesis of Bcl-2 and Bcl-x_L.

DNA Damage

Exposure of cells to radiation or chemotherapeutic agents induces DNA damage, which if severe may trigger apoptotic death. When DNA is damaged, the p53 protein accumulates in cells. It first arrests the cell cycle (at the G_1 phase) to allow the DNA to be repaired before it is replicated (Chapter 5). However, if the damage is too great to be

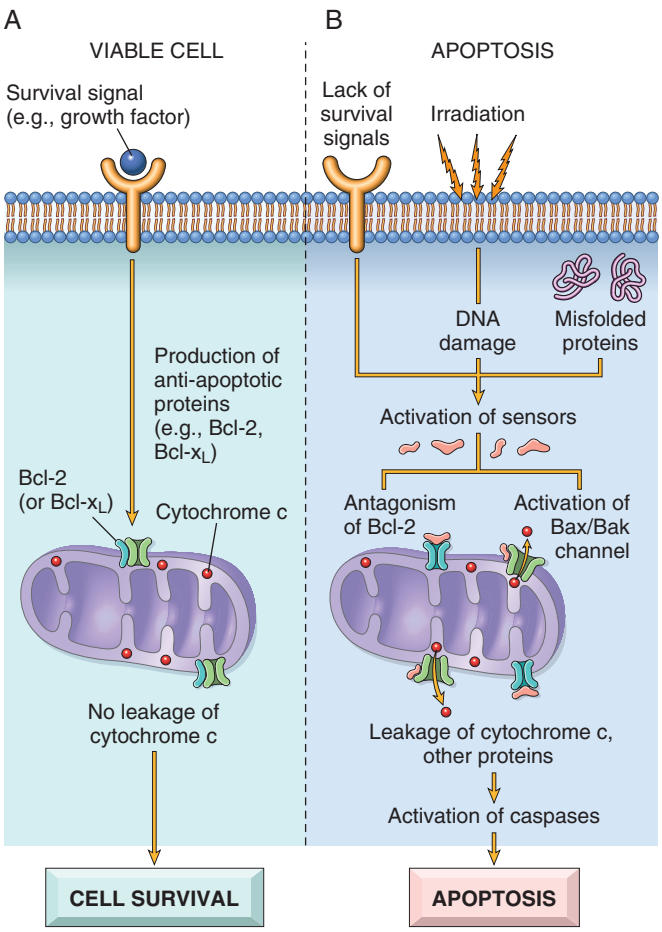


Figure 1–23 The mitochondrial pathway of apoptosis.The induction of apoptosis by the mitochondrial pathway is dependent on a balance between pro- and anti-apoptotic proteins of the Bcl family. The pro-apoptotic proteins include some (sensors) that sense DNA and protein damage and trigger apoptosis and others (effectors) that insert in the mitochondrial membrane and promote leakage of mitochondrial proteins. **A,** In a viable cell, anti-apoptotic members of the Bcl-2 family prevent leakage of mitochondrial proteins. **B,** Various injurious stimuli activate cytoplasmic sensors and lead to reduced production of these anti-apoptotic proteins and increased amounts of pro-apoptotic proteins, resulting in leakage of proteins that are normally sequestered within mitochondria. The mitochondrial proteins that leak out activate a series of caspases, first the initiators and then the executioners, and these enzymes cause fragmentation of the nucleus and ultimately the cell.

repaired successfully, p53 triggers apoptosis, mainly by stimulating sensors that ultimately activate Bax and Bak, and by increasing the synthesis of pro-apoptotic members of the Bcl-2 family. When p53 is mutated or absent (as it is in certain cancers), cells with damaged DNA that would otherwise undergo apoptosis survive. In such cells, the DNA damage may result in mutations or DNA rearrangements (e.g., translocations) that lead to neoplastic transformation (Chapter 5).

Accumulation of Misfolded Proteins: ER Stress

During normal protein synthesis, chaperones in the ER control the proper folding of newly synthesized proteins, and misfolded polypeptides are ubiquitinated and targeted for proteolysis. If, however, unfolded or misfolded proteins accumulate in the ER because of inherited mutations or environmental perturbations, they induce a protective cellular response that is called the *unfolded protein response* (Fig. 1–24). This response activates signaling pathways that increase the production of chaperones and retard protein translation, thus reducing the levels of misfolded proteins in the cell. In circumstances in which the accumulation of misfolded proteins overwhelms these adaptations, the result is *ER stress*, which leads to the activation of caspases and ultimately apoptosis. Intracellular accumulation of abnormally folded proteins, caused by mutations, aging, or unknown environmental factors, may cause diseases by reducing the availability of the normal protein or by inducing cell injury (Table 1–2). Cell death as a result of protein misfolding is now recognized as a feature of a number of neurodegenerative diseases, including Alzheimer, Huntington, and Parkinson diseases, and possibly type 2 diabetes. Deprivation of glucose and oxygen and stresses such as infections also result in protein misfolding, culminating in cell injury and death.

Apoptosis of Self-Reactive Lymphocytes

Lymphocytes capable of recognizing self antigens are normally produced in all individuals. If these lymphocytes encounter self antigens, the cells die by apoptosis. Both the mitochondrial pathway and the Fas death receptor pathway have been implicated in this process (Chapter 4). Failure of apoptosis of self-reactive lymphocytes is one of the causes of autoimmune diseases.

Table 1–2 Diseases Caused by Misfolding of Proteins

Disease	Affected Protein	Pathogenesis
Cystic fibrosis	Cystic fibrosis transmembrane conductance regulator (CFTR)	Loss of CFTR leads to defects in chloride transport
Familial hypercholesterolemia	LDL receptor	Loss of LDL receptor leading to hypercholesterolemia
Tay-Sachs disease	Hexosaminidase β subunit	Lack of the lysosomal enzyme leads to storage of GM_2 gangliosides in neurons
Alpha-1-antitrypsin deficiency	α-1 antitrypsin	Storage of nonfunctional protein in hepatocytes causes apoptosis; absence of enzymatic activity in lungs causes destruction of elastic tissue giving rise to emphysema
Creutzfeld-Jacob disease	Prions	Abnormal folding of PrP^{sc} causes neuronal cell death
Alzheimer disease	A_β peptide	Abnormal folding of A_β peptides causes aggregation within neurons and apoptosis

Shown are selected illustrative examples of diseases in which protein misfolding is thought to be the major mechanism of functional derangement or cell or tissue injury.

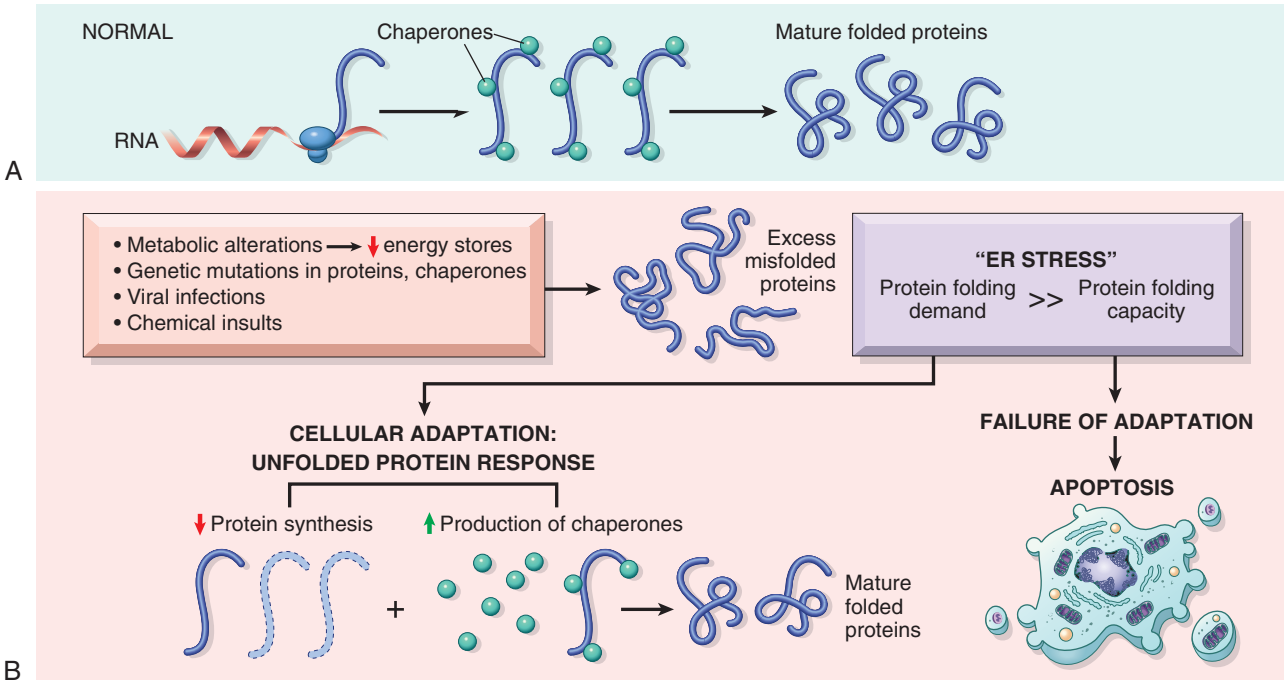


Figure 1–24 The unfolded protein response and ER stress. **A,** In healthy cells, newly synthesized proteins are folded with the help of chaperones and are then incorporated into the cell or secreted. **B,** Various external stresses or mutations induce a state called ER stress, in which the cell is unable to cope with the load of misfolded proteins. Accumulation of these proteins in the ER triggers the unfolded protein response, which tries to restore protein homeostasis; if this response is inadequate, the cell dies by apoptosis.

Cytotoxic T Lymphocyte–Mediated Apoptosis

Cytotoxic T lymphocytes (CTLs) recognize foreign antigens presented on the surface of infected host cells and tumor cells (Chapter 4). On activation, CTL granule proteases called *granzymes* enter the target cells. Granzymes cleave proteins at aspartate residues and are able to activate cellular caspases. In this way, the CTL kills target cells by directly inducing the effector phase of apoptosis, without engaging mitochondria or death receptors. CTLs also express FasL on their surface and may kill target cells by ligation of Fas receptors.

SUMMARY

Apoptosis

- Regulated mechanism of cell death that serves to eliminate unwanted and irreparably damaged cells, with the least possible host reaction
- Characterized by enzymatic degradation of proteins and DNA, initiated by caspases; and by recognition and removal of dead cells by phagocytes
- Initiated by two major pathways:
  - *Mitochondrial (intrinsic) pathway* is triggered by loss of survival signals, DNA damage and accumulation of misfolded proteins (ER stress); associated with leakage of pro-apoptotic proteins from mitochondrial membrane into the cytoplasm, where they trigger caspase activation; inhibited by anti-apoptotic members of the Bcl family, which are induced by survival signals including growth factors.
  - *Death receptor (extrinsic) pathway* is responsible for elimination of self-reactive lymphocytes and damage by cytotoxic T lymphocytes; is initiated by engagement of death receptors (members of the TNF receptor family) by ligands on adjacent cells.

AUTOPHAGY

Autophagy ("self-eating") refers to lysosomal digestion of the cell's own components. It is a survival mechanism in times of nutrient deprivation, such that the starved cell subsists by eating its own contents and recycling these contents to provide nutrients and energy. In this process, intracellular organelles and portions of cytosol are first sequestered within an *autophagic vacuole,* which is postulated to be formed from ribosome-free regions of the ER (Fig. 1–25). The vacuole fuses with lysosomes to form an *autophagolysosome,* in which lysosomal enzymes digest the cellular components. Autophagy is initiated by multiprotein complexes that sense nutrient deprivation and stimulate formation of the autophagic vacuole. With time, the starved cell eventually can no longer cope by devouring itself; at this stage, autophagy may also signal cell death by apoptosis.

Autophagy is also involved in the clearance of misfolded proteins, for instance, in neurons and hepatocytes. Therefore, defective autophagy may be a cause of neuronal death induced by accumulation of these proteins and, subsequently, neurodegenerative diseases. Conversely, pharmacologic activation of autophagy limits the build-up of misfolded proteins in liver cells in animal models,

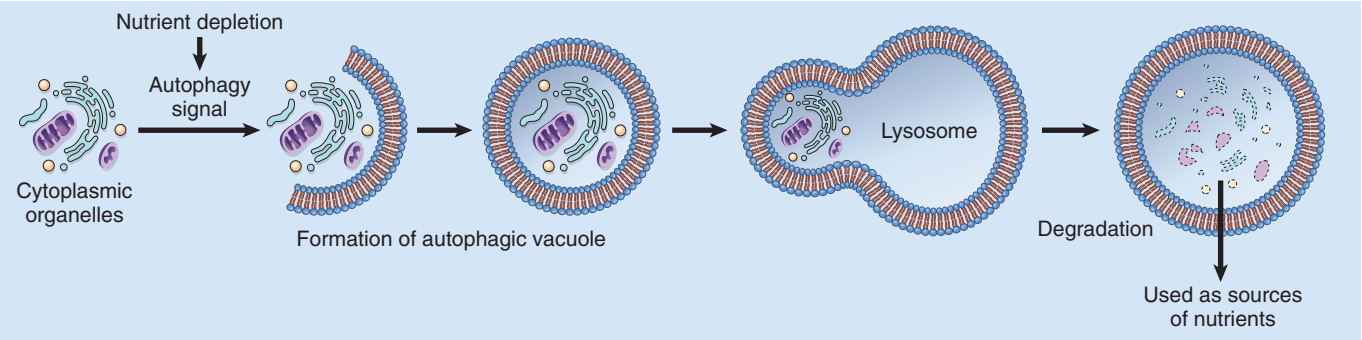


Figure 1–25 Autophagy. Cellular stresses, such as nutrient deprivation, activate autophagy genes (*Atg* genes), which initiate the formation of membrane-bound vesicles in which cellular organelles are sequestered. These vesicles fuse with lysosomes, in which the organelles are digested, and the products are used to provide nutrients for the cell. The same process can trigger apoptosis, by mechanisms that are not well defined.

thereby reducing liver fibrosis. Polymorphisms in a gene involved in autophagy have been associated with inflammatory bowel disease, but the mechanistic link between autophagy and intestinal inflammation is not known. The role of autophagy in cancer is discussed in Chapter 5. Thus, a once little-appreciated survival pathway in cells may prove to have wide-ranging roles in human disease.

We have now concluded the discussion of cell injury and cell death. As we have seen, these processes are the root cause of many common diseases. We end this chapter with brief considerations of three other processes: intracellular accumulations of various substances and extracellular deposition of calcium, both of which are often associated with cell injury, and aging.

INTRACELLULAR ACCUMULATIONS

Under some circumstances cells may accumulate abnormal amounts of various substances, which may be harmless or associated with varying degrees of injury. The substance may be located in the cytoplasm, within organelles (typically lysosomes), or in the nucleus, and it may be synthesized by the affected cells or may be produced elsewhere.

There are four main pathways of abnormal intracellular accumulations (Fig. 1–26):

- Inadequate removal of a normal substance secondary to defects in mechanisms of packaging and transport, as in fatty change in the liver
- Accumulation of an abnormal endogenous substance as a result of genetic or acquired defects in its folding, packaging, transport, or secretion, as with certain mutated forms of α_1-antitrypsin
- Failure to degrade a metabolite due to inherited enzyme deficiencies. The resulting disorders are called *storage diseases* (Chapter 6).
- Deposition and accumulation of an abnormal exogenous substance when the cell has neither the enzymatic machinery to degrade the substance nor the ability to transport it to other sites. Accumulation of carbon or silica particles is an example of this type of alteration.

Fatty Change (Steatosis)

Fatty change refers to any abnormal accumulation of triglycerides within parenchymal cells. It is most often seen in the liver, since this is the major organ involved in fat metabolism, but it may also occur in heart, skeletal muscle, kidney, and other organs. Steatosis may be caused by toxins, protein malnutrition, diabetes mellitus, obesity, or anoxia. *Alcohol abuse and diabetes associated with obesity are the most common causes of fatty change in the liver* (fatty liver) in industrialized nations. This process is discussed in more detail in Chapter 15.

Cholesterol and Cholesteryl Esters

Cellular cholesterol metabolism is tightly regulated to ensure normal cell membrane synthesis without significant intracellular accumulation. However, phagocytic cells may become overloaded with lipid (triglycerides, cholesterol, and cholesteryl esters) in several different pathologic processes. Of these, atherosclerosis is the most important. The role of lipid and cholesterol deposition in the pathogenesis of atherosclerosis is discussed in Chapter 9.

Proteins

Morphologically visible protein accumulations are much less common than lipid accumulations; they may occur when excesses are presented to the cells or if the cells synthesize excessive amounts. In the kidney, for example, trace amounts of albumin filtered through the glomerulus are normally reabsorbed by pinocytosis in the proximal convoluted tubules. However, in disorders with heavy protein leakage across the glomerular filter (e.g., nephrotic syndrome), there is a much larger reabsorption of the protein, and vesicles containing this protein accumulate, giving the histologic appearance of pink, hyaline cytoplasmic droplets. The process is reversible: If the proteinuria abates, the protein droplets are metabolized and disappear. Another example is the marked accumulation of newly synthesized immunoglobulins that may occur in the RER of some plasma cells, forming rounded, eosinophilic *Russell bodies.* Other examples of protein aggregation are discussed elsewhere in this book (e.g., "alcoholic hyaline" in the liver in Chapter 15; neurofibrillary tangles in neurons in Chapter 22).

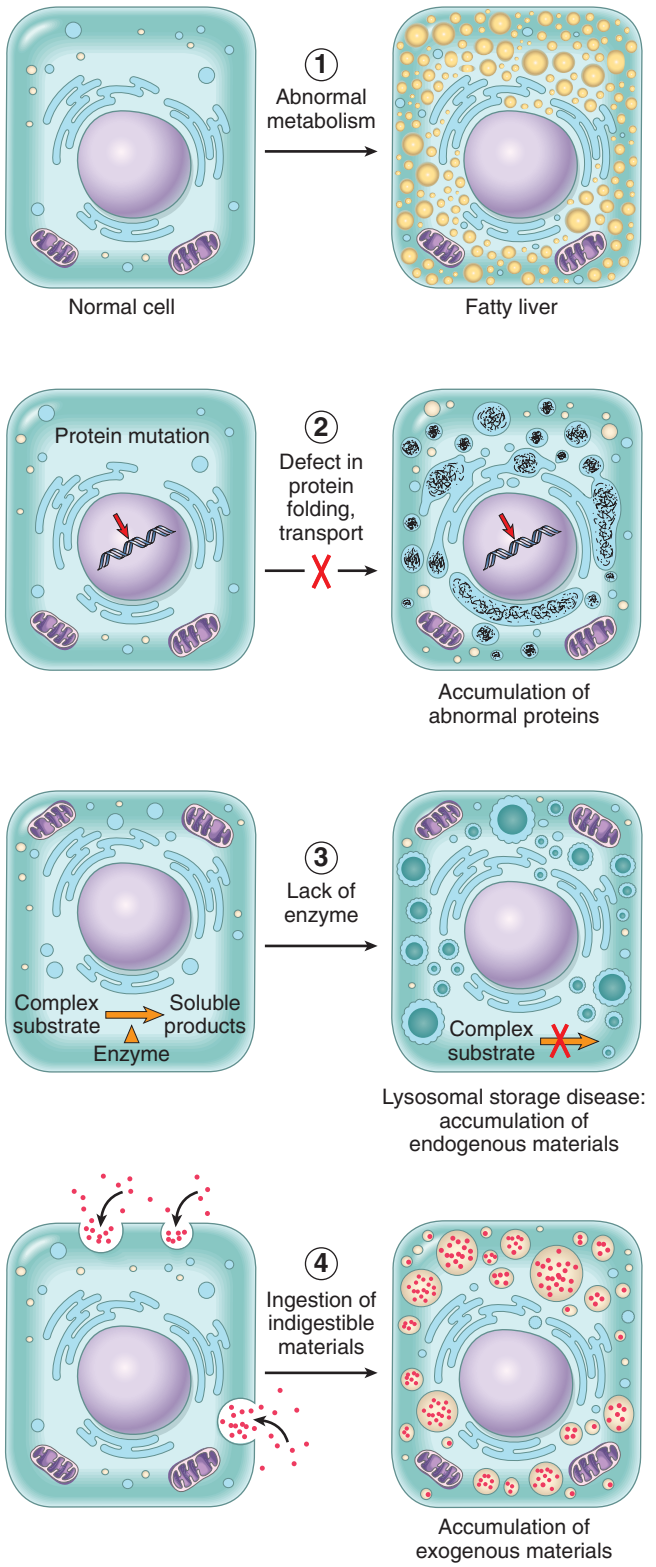


Figure 1–26 Mechanisms of intracellular accumulation: *(1)* Abnormal metabolism, as in fatty change in the liver. *(2)* Mutations causing alterations in protein folding and transport, so that defective molecules accumulate intracellularly. *(3)* A deficiency of critical enzymes responsible for breaking down certain compounds, causing substrates to accumulate in lysosomes, as in lysosomal storage diseases. *(4)* An inability to degrade phagocytosed particles, as in carbon pigment accumulation.

Glycogen

Excessive intracellular deposits of glycogen are associated with abnormalities in the metabolism of either glucose or glycogen. In poorly controlled diabetes mellitus, the prime example of abnormal glucose metabolism, glycogen accumulates in renal tubular epithelium, cardiac myocytes, and β cells of the islets of Langerhans. Glycogen also accumulates within cells in a group of closely related genetic disorders collectively referred to as *glycogen storage diseases,* or *glycogenoses* (Chapter 6).

Pigments

Pigments are colored substances that are either exogenous, coming from outside the body, such as carbon, or endogenous, synthesized within the body itself, such as lipofuscin, melanin, and certain derivatives of hemoglobin.

- The most common exogenous pigment is *carbon* (an example is coal dust), a ubiquitous air pollutant of urban life. When inhaled, it is phagocytosed by alveolar macrophages and transported through lymphatic channels to the regional tracheobronchial lymph nodes. Aggregates of the pigment blacken the draining lymph nodes and pulmonary parenchyma (*anthracosis*) (Chapter 12).
- *Lipofuscin,* or "wear-and-tear pigment," is an insoluble brownish-yellow granular intracellular material that accumulates in a variety of tissues (particularly the heart, liver, and brain) as a function of age or atrophy. Lipofuscin represents complexes of lipid and protein that derive from the free radical–catalyzed peroxidation of polyunsaturated lipids of subcellular membranes. It is not injurious to the cell but is a marker of past free radical injury. The brown pigment (Fig. 1–27), when present in large amounts, imparts an appearance to the tissue that is called *brown atrophy.* By electron microscopy, the pigment appears as perinuclear electron-dense granules (Fig. 1–27, *B*).
- *Melanin* is an endogenous, brown-black pigment that is synthesized by melanocytes located in the epidermis and acts as a screen against harmful ultraviolet radiation. Although melanocytes are the only source of melanin, adjacent basal keratinocytes in the skin can accumulate the pigment (e.g., in freckles), as can dermal macrophages.
- *Hemosiderin* is a hemoglobin-derived granular pigment that is golden yellow to brown and accumulates in tissues when there is a local or systemic excess of iron. Iron is normally stored within cells in association with the protein *apoferritin,* forming ferritin micelles. Hemosiderin pigment represents large aggregates of these ferritin micelles, readily visualized by light and electron microscopy; the iron can be unambiguously identified by the Prussian blue histochemical reaction (Fig. 1–28). Although hemosiderin accumulation is usually pathologic, small amounts of this pigment are normal in the mononuclear phagocytes of the bone marrow, spleen, and liver, where aging red cells are normally degraded. Excessive deposition of hemosiderin, called *hemosiderosis*, and more extensive accumulations of iron seen in *hereditary hemochromatosis*, are described in Chapter 15.

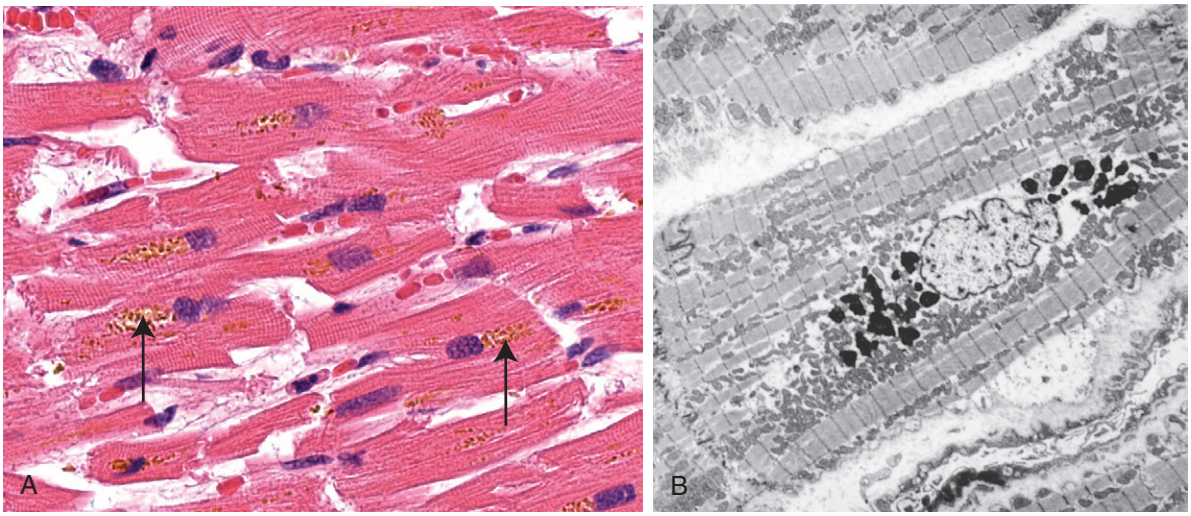


Figure 1–27 Lipofuscin granules in a cardiac myocyte. **A,** Light microscopy (deposits indicated by *arrows*). **B,** Electron microscopy. Note the perinuclear, intralysosomal location.

PATHOLOGIC CALCIFICATION

Pathologic calcification is a common process in a wide variety of disease states; it implies the abnormal deposition of calcium salts, together with smaller amounts of iron, magnesium, and other minerals. When the deposition occurs in dead or dying tissues, it is called *dystrophic calcification; it occurs in the absence of derangements in calcium metabolism* (i.e., with normal serum levels of calcium). In contrast, the deposition of calcium salts in normal tissues is known as *metastatic calcification and is almost always secondary to some derangement in calcium metabolism (hypercalcemia).* Of note, while hypercalcemia is not a prerequisite for dystrophic calcification, it can exacerbate it.

Dystrophic Calcification

Dystrophic calcification is encountered in areas of necrosis of any type. It is virtually inevitable in the *atheromas* of advanced atherosclerosis, associated with intimal injury in the aorta and large arteries and characterized by accumulation of lipids (Chapter 9). Although dystrophic calcification may be an incidental finding indicating insignificant past cell injury, it may also be a cause of organ dysfunction. For example, calcification can develop in aging or damaged heart valves, resulting in severely compromised valve motion. Dystrophic calcification of the aortic valves is an important cause of aortic stenosis in elderly persons (Fig. 10-17, Chapter 10).

The pathogenesis of dystrophic calcification involves *initiation* (or nucleation) and *propagation,* both of which may be either intracellular or extracellular; the ultimate end product is the formation of crystalline *calcium phosphate.* Initiation in extracellular sites occurs in membrane-bound vesicles about 200 nm in diameter; in normal cartilage and bone they are known as *matrix vesicles,* and in pathologic calcification they derive from degenerating cells. It is thought that calcium is initially concentrated in these vesicles by its affinity for membrane phospholipids, while phosphates accumulate as a result of the action of membrane-bound phosphatases. Initiation of intracellular calcification occurs in the mitochondria of dead or dying

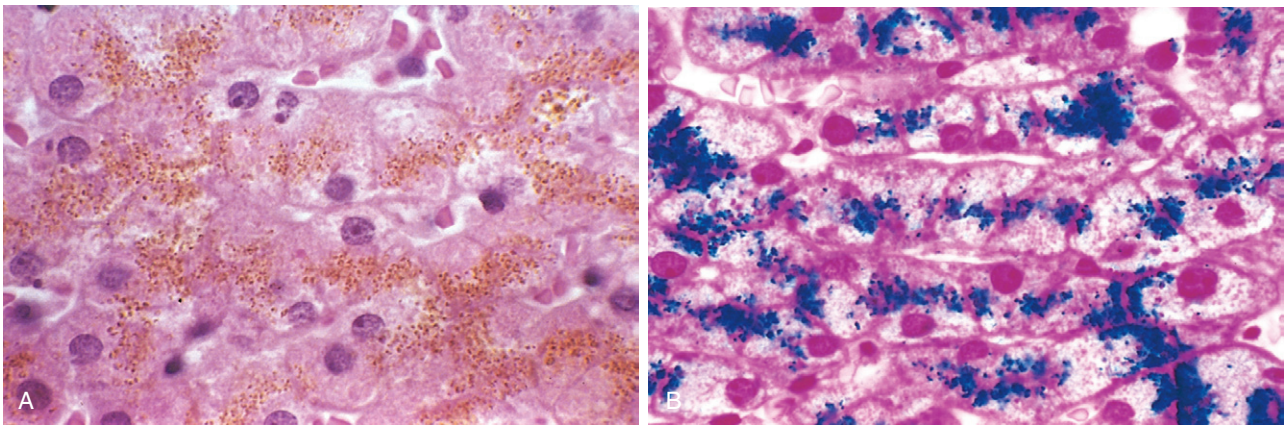


Figure 1–28 Hemosiderin granules in liver cells. **A,** Hematoxylin-eosin–stained section showing golden-brown, finely granular pigment. **B,** Iron deposits revealed by a special staining process called the Prussian blue reaction.

cells that have lost their ability to regulate intracellular calcium. After initiation in either location, propagation of crystal formation occurs. This is dependent on the concentration of Ca^{2+} and PO_4^-, the presence of mineral inhibitors, and the degree of collagenization, which enhances the rate of crystal growth.

Metastatic Calcification

Metastatic calcification can occur in normal tissues whenever there is hypercalcemia. The major causes of hypercalcemia are (1) *increased secretion of parathyroid hormone,* due to either primary parathyroid tumors or production of parathyroid hormone–related protein by other malignant tumors; (2) *destruction of bone* due to the effects of accelerated turnover (e.g., *Paget disease*), immobilization, or tumors (increased bone catabolism associated with multiple myeloma, leukemia, or diffuse skeletal metastases); (3) *vitamin D–related disorders* including vitamin D intoxication and *sarcoidosis* (in which macrophages activate a vitamin D precursor); and (4) *renal failure,* in which phosphate retention leads to *secondary hyperparathyroidism.*

MORPHOLOGY

Regardless of the site, calcium salts are seen on gross examination as fine white granules or clumps, often felt as gritty deposits. Dystrophic calcification is common in areas of caseous necrosis in tuberculosis. Sometimes a tuberculous lymph node is essentially converted to radiopaque stone. On histologic examination, calcification appears as intracellular and/or extracellular basophilic deposits. Over time, heterotopic bone may be formed in the focus of calcification.

Metastatic calcification can occur widely throughout the body but principally affects the interstitial tissues of the vasculature, kidneys, lungs, and gastric mucosa. The calcium deposits morphologically resemble those described in dystrophic calcification. Although they generally do not cause clinical dysfunction, extensive calcifications in the lungs may be evident on radiographs and may produce respiratory deficits, and massive deposits in the kidney **(nephrocalcinosis)** can lead to renal damage.

SUMMARY

Abnormal Intracellular Depositions and Calcifications

Abnormal deposits of materials in cells and tissues are the result of excessive intake or defective transport or catabolism.

- Depositions of *lipids*
  - *Fatty change:* accumulation of free triglycerides in cells, resulting from excessive intake or defective transport (often because of defects in synthesis of transport proteins); manifestation of reversible cell injury
  - *Cholesterol deposition:* result of defective catabolism and excessive intake; in macrophages and smooth muscle cells of vessel walls in atherosclerosis
- Deposition of *proteins:* reabsorbed proteins in kidney tubules; immunoglobulins in plasma cells
- Deposition of *glycogen:* in macrophages of patients with defects in lysosomal enzymes that break down glycogen (glycogen storage diseases)
- Deposition of *pigments:* typically indigestible pigments, such as carbon, lipofuscin (breakdown product of lipid peroxidation), or iron (usually due to overload, as in hemosiderosis)
- Pathologic calcifications
  - *Dystrophic calcification:* deposition of calcium at sites of cell injury and necrosis
  - *Metastatic calcification:* deposition of calcium in normal tissues, caused by hypercalcemia (usually a consequence of parathyroid hormone excess)

CELLULAR AGING

Individuals age because their cells age. Although public attention on the aging process has traditionally focused on its cosmetic manifestations, aging has important health consequences, because age is one of the strongest independent risk factors for many chronic diseases, such as cancer, Alzheimer disease, and ischemic heart disease. Perhaps one of the most striking discoveries about cellular aging is that it is not simply a consequence of cells' "running out of steam," but in fact is regulated by a limited number of genes and signaling pathways that are evolutionarily conserved from yeast to mammals.

Cellular aging is the result of a progressive decline in the life span and functional capacity of cells. Several mechanisms are thought to be responsible for cellular aging (Fig. 1–29):

- *DNA damage.* A variety of metabolic insults that accumulate over time may result in damage to nuclear and mitochondrial DNA. Although most DNA damage is repaired by DNA repair enzymes, some persists and accumulates as cells age. Some aging syndromes are associated with defects in DNA repair mechanisms, and the life span of experimental animals in some models can be increased if responses to DNA damage are enhanced or proteins that stabilize DNA are introduced. A role of free radicals in DNA damage leading to aging has been postulated but remains controversial.
- *Decreased cellular replication.* All normal cells have a limited capacity for replication, and after a fixed number of divisions cells become arrested in a terminally nondividing state, known as *replicative senescence.* Aging is associated with progressive replicative senescence of cells. Cells from children have the capacity to undergo more rounds of replication than do cells from older people. In contrast, cells from patients with *Werner syndrome,* a rare disease characterized by premature aging, have a markedly reduced in vitro life span. In human cells, the mechanism of replicative senescence involves progressive shortening of telomeres, which ultimately results in cell cycle arrest. *Telomeres* are short repeated sequences of DNA present at the ends of linear chromosomes that are important for ensuring the complete replication of chromosome ends and for protecting the ends from fusion and degradation. When somatic cells replicate, a small section of the telomere is not duplicated,

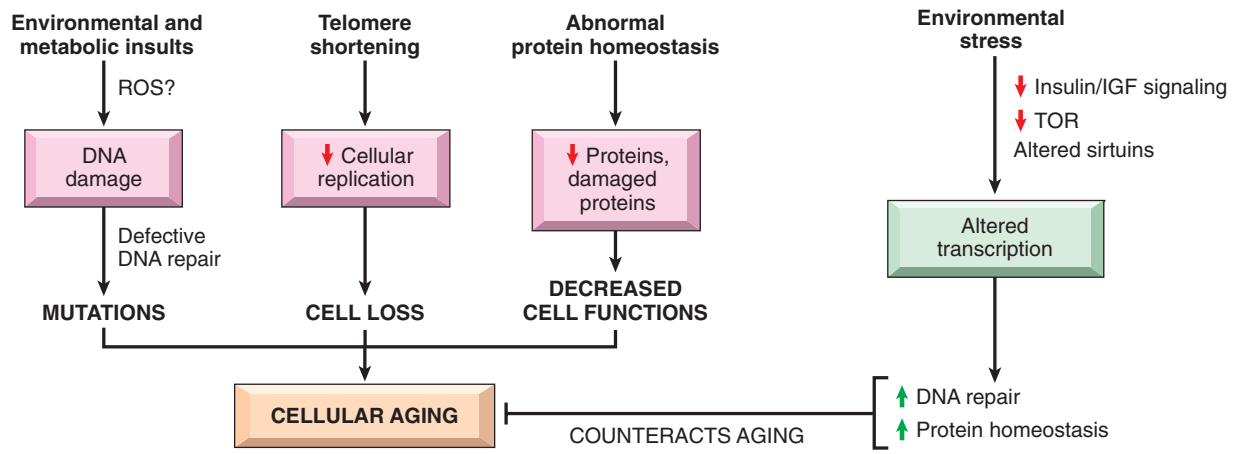


Figure 1–29 Mechanisms that cause and counteract cellular aging. DNA damage, replicative senescence, and decreased and misfolded proteins are among the best described mechanisms of cellular aging. Some environmental stresses, such as calorie restriction, counteract aging by activating various signaling pathways and transcription factors. IGF, insulin-like growth factor; TOR, target of rapamycin.

and telomeres become progressively shortened. As the telomeres become shorter, the ends of chromosomes cannot be protected and are seen as broken DNA, which signals cell cycle arrest. Telomere length is maintained by nucleotide addition mediated by an enzyme called *telomerase.* Telomerase is a specialized RNA-protein complex that uses its own RNA as a template for adding nucleotides to the ends of chromosomes. Telomerase activity is expressed in germ cells and is present at low levels in stem cells, but it is absent in most somatic tissues (Fig. 1–30). Therefore, as most somatic cells age their telomeres become shorter and they exit the cell cycle, resulting in an inability to generate new cells to replace damaged ones. Conversely, in immortalized cancer cells, telomerase is usually reactivated and telomere length is stabilized, allowing the cells to proliferate indefinitely. This is discussed more fully in Chapter 5. Telomere shortening may also decrease the regenerative capacity of stem cells, further contributing to cellular aging. Despite such alluring observations, however, the relationship of telomerase activity and telomere length to aging has yet to be fully established.

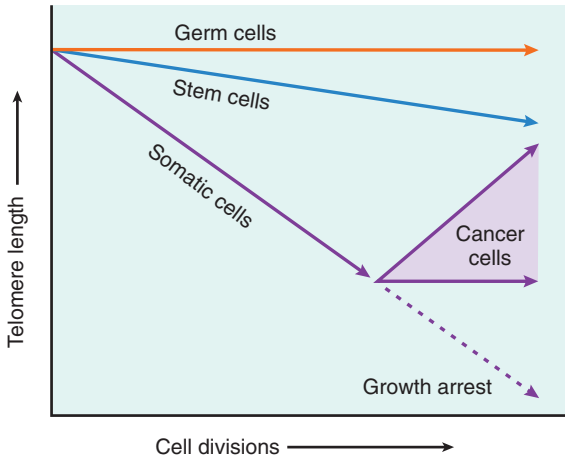


Figure 1–30 The role of telomeres and telomerase in replicative senescence of cells. Telomere length is plotted against the number of cell divisions. In most normal somatic cells there is no telomerase activity, and telomeres progressively shorten with increasing cell divisions until growth arrest, or senescence, occurs. Germ cells and stem cells both contain active telomerase, but only the germ cells have sufficient levels of the enzyme to stabilize telomere length completely. In cancer cells, telomerase is often reactivated.

(Data from Macmillan Publishers Ltd, from Holt SE, et al: Refining the telomer-telomerase hypothesis of aging and cancer. Nat Biotechnol 14:836, 1996.)

- *Defective protein homeostasis.* Over time, cells are unable to maintain normal protein homeostasis, because of increased turnover and decreased synthesis caused by reduced translation of proteins and defective activity of chaperones (which promote normal protein folding), proteasomes (which destroy misfolded proteins) and repair enzymes. Abnormal protein homeostasis can have many effects on cell survival, replication, and functions. In addition, it may lead to accumulation of misfolded proteins, which can trigger pathways of apoptosis.

There has been great interest in defining signaling pathways that counteract the aging process, not only because of their obvious therapeutic potential (the search for the "elixir of youth") but also because elucidating these pathways might tell us about the mechanisms that cause aging. It is now thought that certain *environmental stresses, such as calorie restriction, alter signaling pathways that influence aging* (Fig. 1–29). Among the biochemical alterations that have been described as playing a role in counteracting the aging process are reduced signaling by insulin-like growth factor receptors, reduced activation of kinases (notably the "target of rapamycin," [TOR], and the AKT kinase), and altered transcriptional activity. Ultimately these changes lead to improved DNA repair and protein homeostasis and enhanced immunity, all of which inhibit aging. Environmental stresses may also activate proteins of the Sirtuin family, such as Sir2, which function as protein deacetylases. These proteins may deacetylate and thereby activate DNA repair enzymes, thus stabilizing the DNA; in the absence of these proteins, DNA is more prone to damage. Although the role of sirtuins has received a great deal of attention recently, their importance in the aging process is not yet established.

SUMMARY

Cellular Aging

- Results from combination of accumulating cellular damage (e.g., by free radicals), reduced capacity to divide (replicative senescence), and reduced ability to repair damaged DNA
- *Accumulation of DNA damage:* defective DNA repair mechanisms; conversely DNA repair may be activated by calorie restriction, which is known to prolong aging in model organisms
- *Replicative senescence:* reduced capacity of cells to divide secondary to progressive shortening of chromosomal ends (telomeres)
- *Other factors:* progressive accumulation of metabolic damage; possible roles of growth factors that promote aging in simple model organisms

It should be apparent that the various forms of cellular derangements and adaptations described in this chapter cover a wide spectrum, ranging from adaptations in cell size, growth, and function, to the reversible and irreversible forms of acute cell injury, to the regulated type of cell death represented by apoptosis. Reference is made to these many different alterations throughout this book, because all instances of organ injury and ultimately all cases of clinical disease arise from derangements in cell structure and function.

BIBLIOGRAPHY

Auten RL, Davis JM: Oxygen toxicity and reactive oxygen species: the devil is in the details. Pediatr Res 66:121, 2009. *[A review of the production and degradation of reactive oxygen species, and their roles in cell injury.]*

Balaban RS, Nemoto S, Finkel T: Mitochondria, oxidants, and aging. Cell 120:483, 2005. *[A good review of the role of free radicals in aging.]*

Calado RT, Young NS: Telomere diseases. N Engl J Med 361:2353, 2009. *[An excellent review of the basic biology of telomeres, and how their abnormalities may contribute to cancer, aging, and other diseases.]*

Chipuk JE, Moldoveanu T, Llambl F, et al: The BCL-2 family reunion. Mol Cell 37:299, 2010. *[A review of the biochemistry and biology of the BCL-2 family of apoptosis-regulating proteins.]*

de Groot H, Rauen U: Ischemia-reperfusion injury: processes in pathogenetic networks: a review. Transplant Proc 39:481, 2007. *[A review of the roles of intrinsic cell injury and the inflammatory response in ischemia-reperfusion injury.]*

Dong Z, Saikumar P, Weinberg JM, Venkatachalam MA: Calcium in cell injury and death. Annu Rev Pathol 1:405, 2006. *[A review of the links between calcium and cell injury.]*

Elliott MR, Ravichandran KS: Clearance of apoptotic cells: implications in health and disease. J Cell Biol 189:1059, 2010. *[An excellent review of the mechanisms by which apoptotic cells are cleared, and how abnormalities in these clearance pathways may result in disease.]*

Frey N, Olson EN: Cardiac hypertrophy: the good, the bad, and the ugly. Annu Rev Physiol 65:45, 2003. *[Excellent discussion of the mechanisms of muscle hypertrophy, using the heart as the paradigm.]*

Galluzzi L, Aaronson SA, Abrams J, et al: Guidelines for the use and interpretation of assays for monitoring cell death in higher eukaryotes. Cell Death Differ 16:1093, 2009. *[A practical summary of the morphologic and other techniques for detecting and quantifying dead cells.]*

Haigis MC, Yankner BA: The aging stress response. Mol Cell 40:333, 2010. *[A review of the role of cellular stresses in controlling the aging process.]*

Hotchkiss RS, Strasser A, McDunn JE, Swanson PE: Cell death. N Engl J Med 361:1570, 2009. *[Excellent review of the major pathways of cell death (necrosis, apoptosis, and autophagy-associated death), and their clinical implications and therapeutic targeting.]*

Kenyon CJ: The genetics of ageing. Nature 464:504, 2010. *[An excellent review of the genes that influence aging, based on human genetic syndromes and studies with mutant model organisms.]*

Kroemer G, Marino G, Levine B: Autophagy and the integrated stress response. Mol Cell 40:280, 2010. *[An excellent discussion of the biology, biochemical pathways, and physiologic roles of autophagy.]*

Kundu M, Thompson CB: Autophagy: basic principles and relevance to disease. Annu Rev Pathol 3:427, 2008. *[A discussion of the biology of autophagy and its potential contribution to a variety of disease states.]*

Lin JH, Walter P, Yen TSB: Endoplasmic reticulum stress in disease pathogenesis. Annu Rev Pathol 3:399, 2008. *[A review of the biology and disease relevance of the unfolded protein response and ER stress induced by unfolded proteins.]*

Lombard DB, Chua KF, Mostoslavsky R, et al: DNA repair, genome stability, and aging. Cell 120:497, 2005. *[The role of DNA damage in cellular aging.]*

McKinnell IW, Rudnicki MA: Molecular mechanisms of muscle atrophy. Cell 119:907, 2004. *[Discussion of the mechanisms of cellular atrophy.]*

Newmeyer DD, Ferguson-Miller S: Mitochondria: releasing power for life and unleashing the machineries of death. Cell 112:481, 2003. *[Excellent review of the many functions of mitochondria, with an emphasis on their role in cell death.]*

Sahin E, DePinho RA: Linking functional decline of telomeres, mitochondria and stem cells during ageing. Nature 464:520, 2010. *[An excellent review of stem cell abnormalities that contribute to aging.]*

Tosh D, Slack JM: How cells change their phenotype. Nat Rev Mol Cell Biol 3:187, 2002. *[Review of metaplasia and the roles of stem cells and genetic reprogramming.]*

Valko M, Leibfritz D, Moncol J, et al: Free radicals and antioxidants in normal physiological functions and human disease. Int J Biochem Cell Biol 39:44, 2007. *[An interesting discussion of the biochemistry of reactive oxygen and nitrogen-derived free radicals, their roles in cell injury, and their physiologic functions as signaling molecules.]*

See Targeted Therapy available online at studentconsult.com

CHAPTER 2

Inflammation and Repair

CHAPTER CONTENTS

OVERVIEW OF INFLAMMATION AND TISSUE REPAIR

The survival of all organisms requires that they eliminate foreign invaders, such as infectious agents, and damaged tissues. These functions are mediated by a complex host response called *inflammation. Inflammation is a protective response involving host cells, blood vessels, and proteins and other mediators that is intended to eliminate the initial cause of cell injury, as well as the necrotic cells and tissues resulting from the original insult, and to initiate the process of repair.* Inflammation accomplishes its protective mission by first diluting, destroying, or otherwise neutralizing harmful agents (e.g., microbes, toxins). It then sets into motion the events that eventually heal and repair the sites of injury. Without inflammation, infections would go unchecked and wounds would never heal. In the context of infections, inflammation is one component of a protective response that immunologists refer to as innate immunity (Chapter 4).

Although inflammation helps clear infections and other noxious stimuli and initiates repair, the inflammatory reaction and the subsequent repair process can themselves cause considerable harm. The components of the inflammatory reaction that destroy and eliminate microbes and dead tissues are also capable of injuring normal tissues. Therefore, injury may accompany entirely normal, beneficial inflammatory reactions, and the damage may even become the dominant feature if the reaction is very strong (e.g., when the infection is severe), prolonged (e.g., when the eliciting agent resists eradication), or inappropriate (e.g., when it is directed against self-antigens in autoimmune diseases, or against usually harmless environmental antigens (e.g., in allergic disorders). Some of the most vexing diseases of humans are disorders that result from inappropriate, often chronic, inflammation. Thus, the process of inflammation is fundamental to virtually all of clinical medicine.

The cells and molecules of host defense, including leukocytes and plasma proteins, normally circulate in the blood, and the goal of the inflammatory reaction is to bring them to the site of infection or tissue damage. In addition, resident cells of vascular walls and the cells and proteins of the extracellular matrix (ECM) are also involved in inflammation and repair (Fig. 2–1). Before we describe the process of inflammation in detail, some of the basic features will be highlighted.

Inflammation can be acute or chronic (Table 2–1). Acute inflammation is rapid in onset and of short duration, lasting

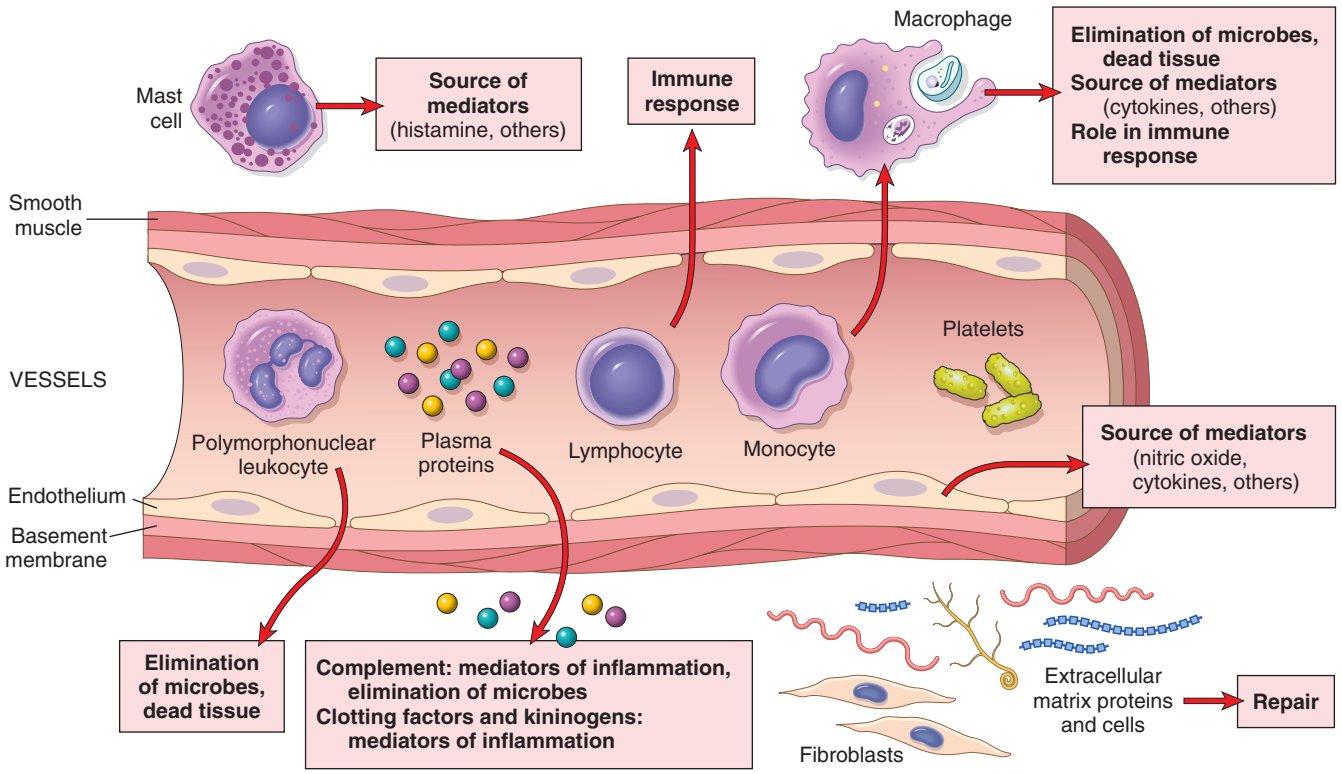


Figure 2–1 The components of acute and chronic inflammatory responses and their principal functions. The roles of these cells and molecules in inflammation are described in this chapter.

from a few minutes to as long as a few days, and is characterized by fluid and plasma protein exudation and a predominantly neutrophilic leukocyte accumulation. Chronic inflammation may be more insidious, is of longer duration (days to years), and is typified by influx of lymphocytes and macrophages with associated vascular proliferation and fibrosis (scarring). As we shall see later, however, these two basic forms of inflammation may coexist, and many variables modify their course and histologic appearance.

Inflammation is induced by chemical mediators that are produced by host cells in response to injurious stimuli. When a microbe enters a tissue or the tissue is injured, the presence of the infection or damage is sensed by resident cells, mainly macrophages, but also dendritic cells, mast cells, and other cell types. These cells secrete molecules (cytokines and other mediators) that induce and regulate the subsequent inflammatory response. Inflammatory mediators are also produced from plasma proteins that react with the microbes or to injured tissues. Some of these mediators promote the efflux of plasma and the recruitment of circulating leukocytes to the site where the offending agent is located. The recruited leukocytes are activated and they try to remove the offending agent by phagocytosis. An unfortunate side effect of the activation of leukocytes may be damage to normal host tissues.

The external manifestations of inflammation, often called its cardinal signs, are heat (calor), redness (rubor), swelling (tumor), pain (dolor), and loss of function (functio laesa). The first four of these were described more than 2000 years ago by a Roman encyclopedist named Celsus, who wrote the then-famous text *De Medicina,* and the fifth was added in the late 19th century by Rudolf Virchow, known as the "father of modern pathology." These manifestations occur as consequences of the vascular changes and leukocyte recruitment and activation, as will be evident from the discussion that follows.

Inflammation is normally controlled and self-limited. The mediators and cells are activated only in response to the injurious stimulus and are short-lived, and they are degraded or become inactive as the injurious agent is eliminated. In addition, various anti-inflammatory mechanisms become active. If the injurious agent cannot be quickly eliminated, the result may be chronic inflammation, which can have serious pathologic consequences.

Table 2–1 Features of Acute and Chronic Inflammation

Feature	Acute	Chronic
Onset	Fast: minutes or hours	Slow: days
Cellular infiltrate	Mainly neutrophils	Monocytes/macrophages and lymphocytes
Tissue injury, fibrosis	Usually mild and self-limited	Often severe and progressive
Local and systemic signs	Prominent	Less prominent; may be subtle

SUMMARY

General Features of Inflammation

- Inflammation is a defensive host response to foreign invaders and necrotic tissue, but it is itself capable of causing tissue damage.
- The main components of inflammation are a vascular reaction and a cellular response; both are activated by mediators derived from plasma proteins and various cells.
- The steps of the inflammatory response can be remembered as the five *R*s: (1) *r*ecognition of the injurious agent, (2) *r*ecruitment of leukocytes, (3) *r*emoval of the agent, (4) *r*egulation (control) of the response, and (5) *r*esolution (repair).
- The outcome of acute inflammation is either elimination of the noxious stimulus, followed by decline of the reaction and repair of the damaged tissue, or persistent injury resulting in chronic inflammation.

ACUTE INFLAMMATION

The acute inflammatory response rapidly delivers leukocytes and plasma proteins to sites of injury. Once there, leukocytes clear the invaders and begin the process of digesting and getting rid of necrotic tissues.

Acute inflammation has two major components (Fig. 2–2):

- *Vascular changes:* alterations in vessel caliber resulting in increased blood flow (vasodilation) and changes in the vessel wall that permit plasma proteins to leave the circulation (increased vascular permeability). In addition, endothelial cells are activated, resulting in increased adhesion of leukocytes and migration of the leukocytes through the vessel wall.
- *Cellular events:* emigration of the leukocytes from the circulation and accumulation in the focus of injury (cellular recruitment), followed by activation of the leukocytes, enabling them to eliminate the offending agent. The principal leukocytes in acute inflammation are neutrophils (polymorphonuclear leukocytes).

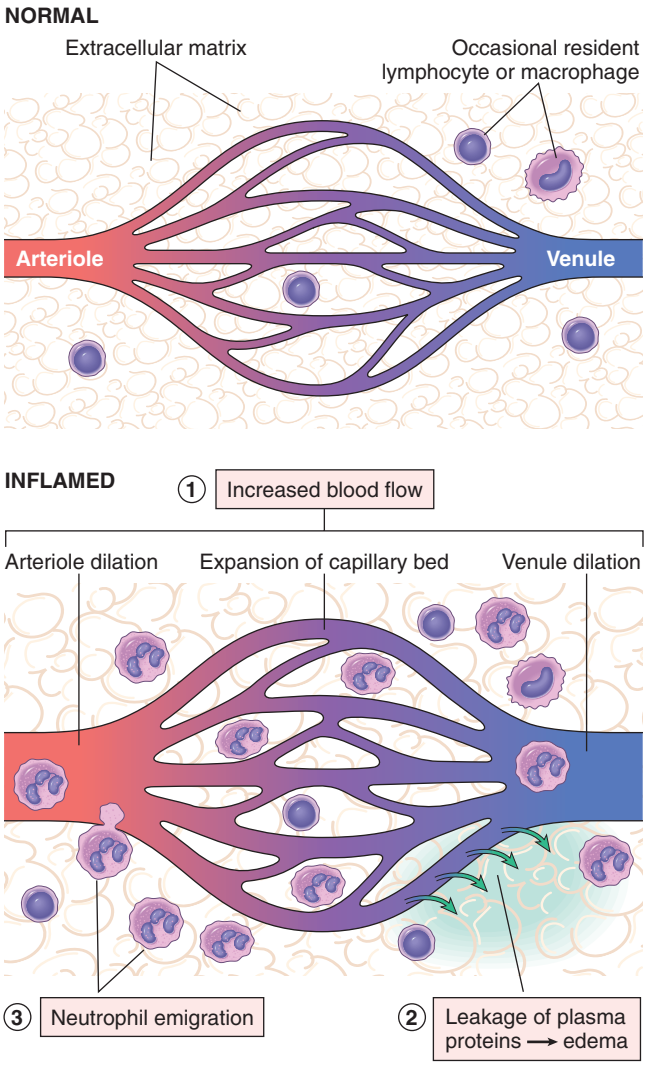


Figure 2–2 Vascular and cellular reactions of acute inflammation. The major local manifestations of acute inflammation, compared with normal, are *(1)* vascular dilation and increased blood flow (causing erythema and warmth), *(2)* extravasation of plasma fluid and proteins (edema), and *(3)* leukocyte (mainly neutrophil) emigration and accumulation.

Stimuli for Acute Inflammation

Acute inflammatory reactions may be triggered by a variety of stimuli:

- *Infections* (bacterial, viral, fungal, parasitic) are among the most common and medically important causes of inflammation.
- *Trauma* (blunt and penetrating) and various physical and chemical agents (e.g., thermal injury, such as burns or frostbite; irradiation; toxicity from certain environmental chemicals) injure host cells and elicit inflammatory reactions.
- *Tissue necrosis* (from any cause), including ischemia (as in a myocardial infarct) and physical and chemical injury
- *Foreign bodies* (splinters, dirt, sutures, crystal deposits)
- *Immune reactions* (also called *hypersensitivity reactions*) against environmental substances or against "self" tissues. Because the stimuli for these inflammatory responses often cannot be eliminated or avoided, such reactions tend to persist, with features of chronic inflammation. The term "immune-mediated inflammatory disease" is sometimes used to refer to this group of disorders.

Although each of these stimuli may induce reactions with some distinctive characteristics, in general, all inflammatory reactions have the same basic features.

In this section, we describe first how inflammatory stimuli are recognized by the host, then the typical reactions of acute inflammation and its morphologic features, and finally the chemical mediators responsible for these reactions.

Recognition of Microbes, Necrotic Cells, and Foreign Substances

A fundamental question relating to activation of the host response is how cells recognize the presence of potentially harmful agents such as microbes in the tissues. It was postulated that microbes and dead cells must elicit some sort of "danger signals" that distinguish them from normal tissues and mobilize the host response. It is now established that *phagocytes, dendritic cells (cells in connective tissue and organs that capture microbes and initiate responses to them), and many other cells, such as epithelial cells, express receptors that are designed to sense the presence of infectious pathogens and substances released from dead cells.* These receptors have been called "pattern recognition receptors" because they recognize structures (i.e., molecular patterns) that are common to many microbes or to dead cells. The two most important families of these receptors are the following:

- *Toll-like receptors (TLRs)* are microbial sensors that are named for the founding member called *Toll,* which was discovered in *Drosophila.* There are ten mammalian TLRs, which recognize products of bacteria (such as endotoxin and bacterial DNA), viruses (such as double-stranded RNA), and other pathogens (Fig. 2–3, *A*). TLRs are located in plasma membranes and endosomes, so they are able to detect extracellular and ingested microbes. They are complemented by cytoplasmic and membrane molecules, from several other families, that also recognize microbial products. TLRs and the other receptors recognize products of different types of microbes and thus provide defense against essentially all classes of infectious pathogens. Recognition of microbes by these receptors activates transcription factors that stimulate the production of a number of secreted and membrane proteins. These proteins include mediators of inflammation, antiviral cytokines (interferons), and proteins that promote lymphocyte activation and even more potent immune responses. We return to TLRs in Chapter 4, when we discuss innate immunity, the early defense against infections.
- The *inflammasome* is a multi-protein cytoplasmic complex that recognizes products of dead cells, such as uric acid and extracellular ATP, as well as crystals and some microbial products. Triggering of the inflammasome results in activation of an enzyme called caspase-1, which cleaves precursor forms of the inflammatory

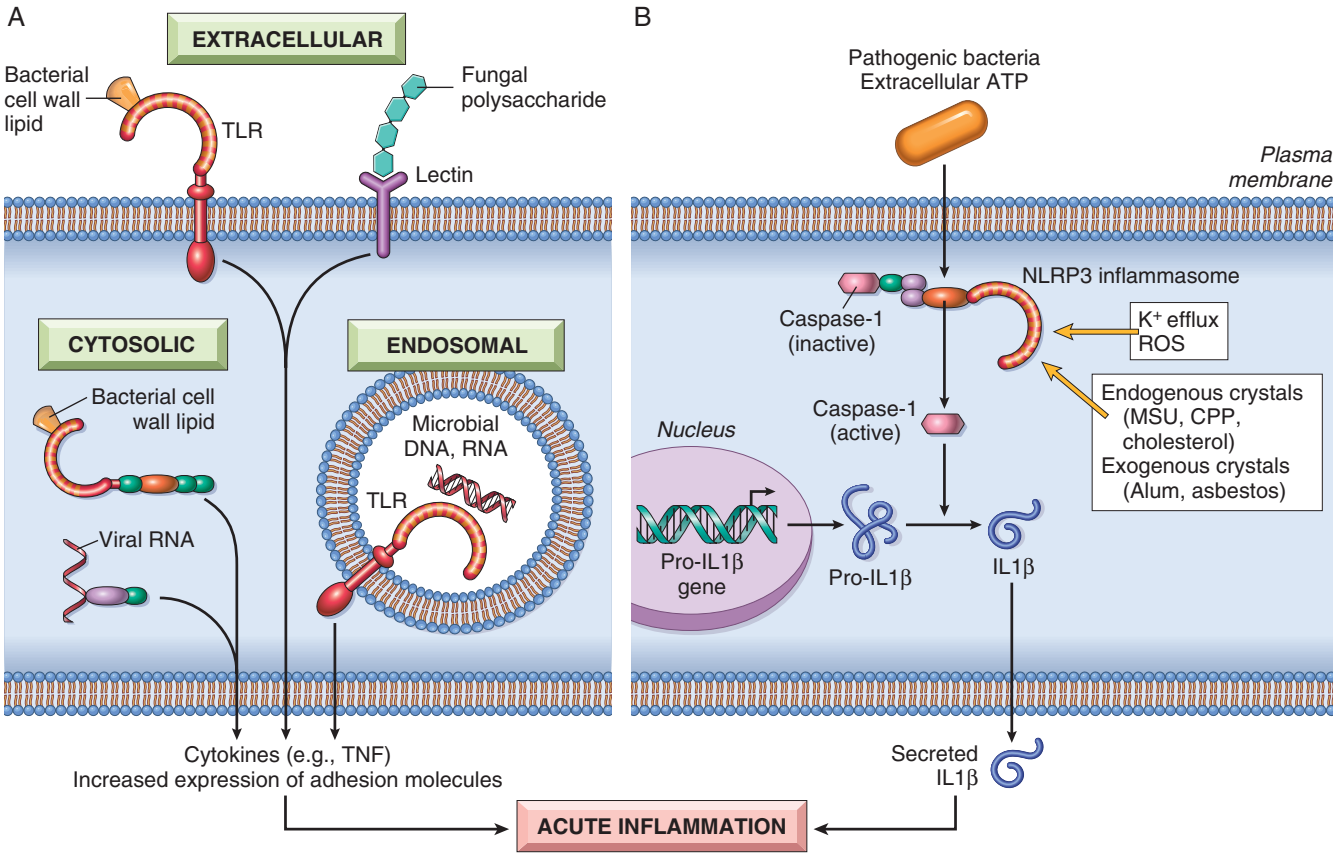


Figure 2–3 Sensors of microbes and dead cells: Phagocytes, dendritic cells, and many types of epithelial cells express different classes of receptors that sense the presence of microbes and dead cells. **A, Toll-like receptors** (TLRs) located in the plasma membrane and endosomes and other cytoplasmic and plasma membrane receptors (members of families other than TLRs) recognize products of different classes of microbes. The proteins produced by TLR activation have numerous functions; only their role in inflammation is shown. **B,** The **inflammasome** is a protein complex that recognizes products of dead cells and some microbes and induces the secretion of biologically active interleukin-1 (IL-1). The inflammasome consists of a sensor protein (a leucine-rich protein called NLRP3), an adaptor, and the enzyme caspase-1, which is converted from an inactive to an active form. (Note that the inflammasome is distinct from phagolysosomes, which also are present in the cytoplasm but are vesicles that serve different functions in inflammation, as discussed later in the chapter.) CPP, calcium pyrophosphate; MSU, monosodium urate.

cytokine interleukin-1β (IL-1β) into its biologically active form (Fig. 2–3, *B*). As discussed later, IL-1 is an important mediator of leukocyte recruitment in the acute inflammatory response, and the leukocytes phagocytose and destroy dead cells. The joint disease, gout, is caused by deposition of urate crystals, which are ingested by phagocytes and activate the inflammasome, resulting in IL-1 production and acute inflammation. IL-1 antagonists are effective treatments in cases of gout that are resistant to conventional anti-inflammatory therapy. Recent studies have shown that cholesterol crystals and free fatty acids also activate the inflammasome, suggesting that IL-1 plays a role in common diseases such as atherosclerosis (associated with deposition of cholesterol crystals in vessel walls) and obesity-associated type 2 diabetes. This finding raises the possibility of treating these diseases by blocking IL-1.

The functions of these sensors are referred to throughout the chapter. We now proceed with a discussion of the principal reactions of acute inflammation.

Vascular Changes

The main vascular reactions of acute inflammation are increased blood flow secondary to vasodilation and increased vascular permeability, both designed to bring blood cells and proteins to sites of infection or injury. While the initial encounter of an injurious stimulus, such as a microbe, is with macrophages and other cells in the connective tissue, the vascular reactions triggered by these interactions soon follow and dominate the early phase of the response.

Changes in Vascular Caliber and Flow

Changes in blood vessels are initiated rapidly after infection or injury but evolve at variable rates, depending on the nature and severity of the original inflammatory stimulus.

- After transient vasoconstriction (lasting only for seconds), arteriolar vasodilation occurs, resulting in locally increased blood flow and engorgement of the down-stream capillary beds (Fig. 2–2). This vascular expansion is the cause of the redness (*erythema*) and warmth characteristic of acute inflammation, and mentioned previously as two of the cardinal signs of inflammation.
- The microvasculature becomes more permeable, and protein-rich fluid moves into the extravascular tissues. This causes the red cells in the flowing blood to become more concentrated, thereby increasing blood viscosity and slowing the circulation. These changes are reflected microscopically by numerous dilated small vessels packed with red blood cells, called *stasis.*
- As stasis develops, leukocytes (principally neutrophils) begin to accumulate along the vascular endothelial surface—a process called *margination.* This is the first step in the journey of the leukocytes through the vascular wall into the interstitial tissue (described later).

Increased Vascular Permeability

Increasing vascular permeability leads to the movement of protein-rich fluid and even blood cells into the extravascular tissues (Fig. 2–4). This in turn increases the osmotic pressure of the interstitial fluid, leading to more outflow of

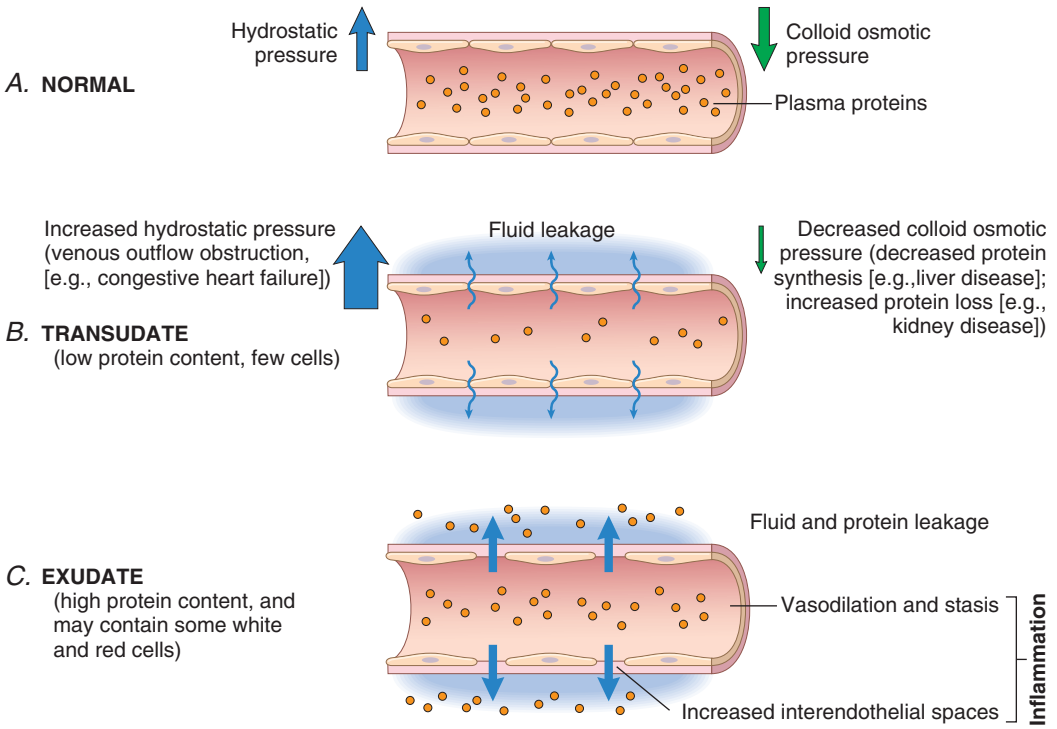


Figure 2–4 Formation of transudates and exudates. *A*, Normal hydrostatic pressure (*blue arrows*) is approximately 32 mm Hg at the arterial end of a capillary bed and 12 mm Hg at the venous end; the mean colloid osmotic pressure of tissues is approximately 25 mm Hg (*green arrows*), which is nearly equal to the mean capillary pressure. Therefore, the net flow of fluid across the vascular bed is almost nil. *B*, A transudate is formed when fluid leaks out because of increased hydrostatic pressure or decreased osmotic pressure. *C*, An exudate is formed in inflammation because vascular permeability increases as a result of the increase in interendothelial spaces.

water from the blood into the tissues. The resulting protein-rich fluid accumulation is called an *exudate.* Exudates must be distinguished from *transudates,* which are interstitial fluid accumulations caused by increased hydrostatic pressure, usually a consequence of reduced venous return. Transudates typically contain low concentrations of protein and few or no blood cells. Fluid accumulation in extravascular spaces, whether from an exudate or a transudate, produces tissue *edema.* Whereas exudates are typical of inflammation, transudates accumulate in various noninflammatory conditions, which are mentioned in Figure 2–4 and described in more detail in Chapter 3.

Several mechanisms may contribute to increased vascular permeability in acute inflammatory reactions:

- *Endothelial cell contraction leading to intercellular gaps in postcapillary venules* is the most common cause of increased vascular permeability. Endothelial cell contraction occurs rapidly after binding of histamine, bradykinin, leukotrienes, and many other mediators to specific receptors, and is usually short-lived (15 to 30 minutes). A slower and more prolonged retraction of endothelial cells, resulting from changes in the cytoskeleton, may be induced by cytokines such as tumor necrosis factor (TNF) and interleukin-1 (IL-1). This reaction may take 4 to 6 hours to develop after the initial trigger and persist for 24 hours or more.
- *Endothelial injury* results in vascular leakage by causing endothelial cell necrosis and detachment. Endothelial cells are damaged after severe injury such as with burns and some infections. In most cases, leakage begins immediately after the injury and persists for several hours (or days) until the damaged vessels are thrombosed or repaired. Venules, capillaries, and arterioles can all be affected, depending on the site of the injury. Direct injury to endothelial cells may also induce a delayed prolonged leakage that begins after a delay of 2 to 12 hours, lasts for several hours or even days, and involves venules and capillaries. Examples are mild to moderate thermal injury, certain bacterial toxins, and x- or ultraviolet irradiation (i.e., the sunburn that has spoiled many an evening after a day in the sun). Endothelial cells may also be damaged as a consequence of leukocyte accumulation along the vessel wall. Activated leukocytes release many toxic mediators, discussed later, that may cause endothelial injury or detachment.
- *Increased transcytosis* of proteins by way of an intracellular vesicular pathway augments venular permeability, especially after exposure to certain mediators such as vascular endothelial growth factor (VEGF). Transcytosis occurs through channels formed by fusion of intracellular vesicles.
- *Leakage from new blood vessels.* As described later, tissue repair involves new blood vessel formation (angiogenesis). These vessel sprouts remain leaky until proliferating endothelial cells mature sufficiently to form intercellular junctions. New endothelial cells also have increased expression of receptors for vasoactive mediators, and some of the factors that stimulate angiogenesis (e.g., VEGF) also directly induce increased vascular permeability.

Although these mechanisms of vascular permeability are separable, all of them may participate in the response to a particular stimulus. For example, in a thermal burn, leakage results from chemically mediated endothelial contraction, as well as from direct injury and leukocyte-mediated endothelial damage.

Responses of Lymphatic Vessels

In addition to blood vessels, lymphatic vessels also participate in the inflammatory response. In inflammation, lymph flow is increased and helps drain edema fluid, leukocytes, and cell debris from the extravascular space. In severe inflammatory reactions, especially to microbes, the lymphatics may transport the offending agent, contributing to its dissemination. The lymphatics may become secondarily inflamed (*lymphangitis*), as may the draining lymph nodes (*lymphadenitis*). Inflamed lymph nodes are often enlarged because of hyperplasia of the lymphoid follicles and increased numbers of lymphocytes and phagocytic cells lining the sinuses of the lymph nodes. This constellation of pathologic changes is termed reactive, or inflammatory, lymphadenitis (Chapter 11). For clinicians, the presence of red streaks near a skin wound is a telltale sign of an infection in the wound. This streaking follows the course of the lymphatic channels and is diagnostic of lymphangitis; it may be accompanied by painful enlargement of the draining lymph nodes, indicating lymphadenitis.

SUMMARY

Vascular Reactions in Acute Inflammation

- Vasodilation is induced by chemical mediators such as histamine (described later) and is the cause of erythema and stasis of blood flow.
- Increased vascular permeability is induced by histamine, kinins, and other mediators that produce gaps between endothelial cells; by direct or leukocyte-induced endothelial injury; and by increased passage of fluids through the endothelium. This increased permeability allows plasma proteins and leukocytes to enter sites of infection or tissue damage; fluid leak through blood vessels results in edema.

Cellular Events: Leukocyte Recruitment and Activation

As mentioned earlier, an important function of the inflammatory response is to deliver leukocytes to the site of injury and to activate them. Leukocytes ingest offending agents, kill bacteria and other microbes, and eliminate necrotic tissue and foreign substances. A price that is paid for the defensive potency of leukocytes is that once activated, they may induce tissue damage and prolong inflammation, since the leukocyte products that destroy microbes can also injure normal host tissues. Therefore, host defense mechanisms include checks and balances that ensure that leukocytes are recruited and activated only when and where they are needed (i.e., in response to foreign invaders and dead tissues). Systemic activation of leukocytes can, in fact, have detrimental consequences, as in septic shock (Chapter 3).

Leukocyte Recruitment

Leukocytes normally flow rapidly in the blood, and in inflammation, they have to be stopped and brought to the offending agent or the site of tissue damage, which are typically outside the vessels. The sequence of events in the recruitment of leukocytes from the vascular lumen to the extravascular space consists of (1) margination and rolling along the vessel wall; (2) firm adhesion to the endothelium; (3) transmigration between endothelial cells; and (4) migration in interstitial tissues toward a chemotactic stimulus (Fig. 2–5). Rolling, adhesion, and transmigration are mediated by the interactions of adhesion molecules on leukocytes and endothelial surfaces (see later on). Chemical mediators—chemoattractants and certain cytokines—affect these processes by modulating the surface expression and binding affinity of the adhesion molecules and by stimulating directional movement of the leukocytes.

Margination and Rolling. As blood flows from capillaries into postcapillary venules, circulating cells are swept by laminar flow against the vessel wall. Because the smaller red cells tend to move faster than the larger white cells, leukocytes are pushed out of the central axial column and thus have a better opportunity to interact with lining endothelial cells, especially as stasis sets in. This process of leukocyte accumulation at the periphery of vessels is called *margination.* If the endothelial cells are activated by cytokines and other mediators produced locally, they express adhesion molecules to which the leukocytes attach loosely. These cells bind and detach and thus begin to tumble on the endothelial surface, a process called *rolling.*

The weak and transient interactions involved in rolling are mediated by the *selectin* family of adhesion molecules (Table 2–2). Selectins are receptors expressed on leukocytes and endothelium that contain an extracellular domain that binds sugars (hence the lectin part of the name). The three members of this family are E-selectin (also called CD62E), expressed on endothelial cells; P-selectin (CD62P), present on platelets and endothelium; and L-selectin (CD62L), on the surface of most leukocytes. Selectins bind sialylated oligosaccharides (e.g., sialyl–Lewis X on leukocytes) that are attached to mucin-like glycoproteins on various cells. The endothelial selectins are typically expressed at low levels or are not present at all on unactivated endothelium, and are up-regulated after stimulation by cytokines and other mediators. Therefore, binding of leukocytes is largely restricted to endothelium at sites of infection or tissue injury (where the mediators are produced). For example, in unactivated endothelial cells, P-selectin is found primarily in intracellular Weibel-Palade bodies; however, within minutes of exposure to mediators such as histamine or thrombin, P-selectin is distributed to the cell surface, where it can facilitate leukocyte binding. Similarly, E-selectin and the ligand for L-selectin, which are not expressed on normal endothelium, are induced after stimulation by the cytokines IL-1 and TNF.

Adhesion. The rolling leukocytes are able to sense changes in the endothelium that initiate the next step in the reaction of leukocytes, which is firm *adhesion* to endothelial surfaces. This adhesion is mediated by *integrins* expressed on leukocyte cell surfaces interacting with their ligands on endothelial cells (Fig. 2–5 and Table 2–2). Integrins are

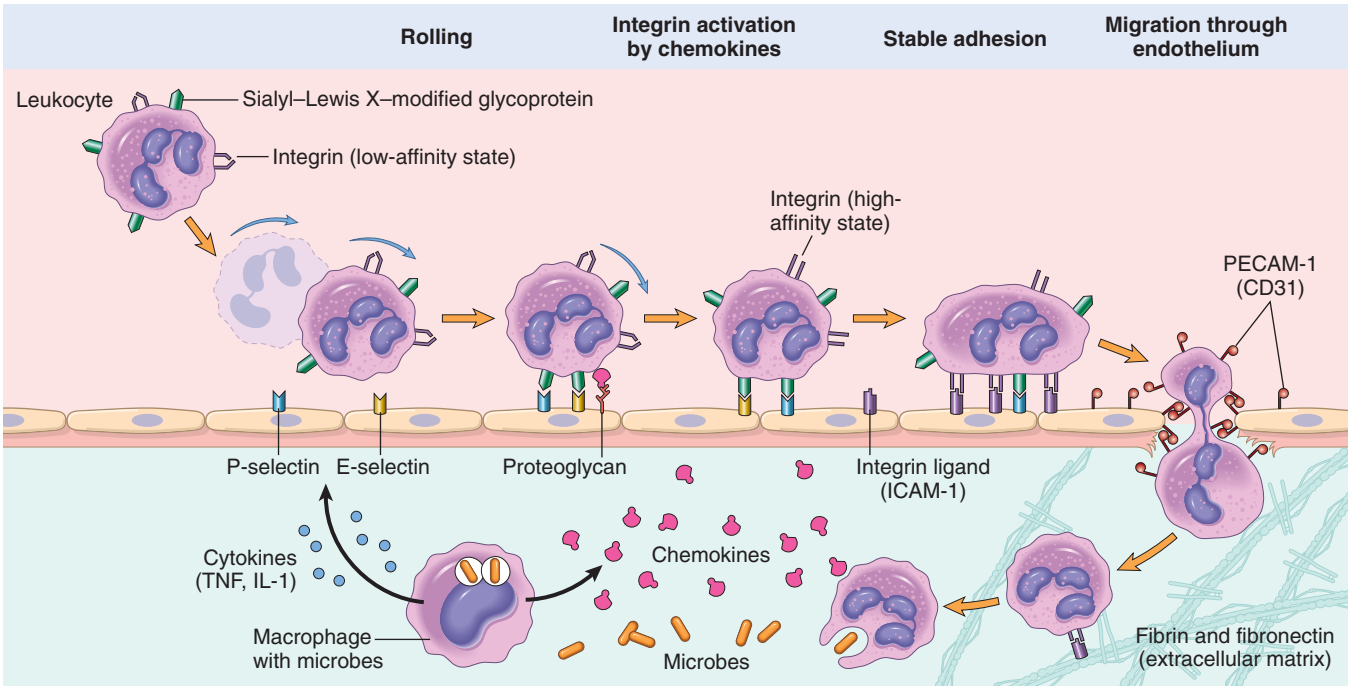


Figure 2–5 Mechanisms of leukocyte migration through blood vessels. The leukocytes (neutrophils shown here) first roll, then become activated and adhere to endothelium, then transmigrate across the endothelium, pierce the basement membrane, and migrate toward chemoattractants emanating from the source of injury. Different molecules play predominant roles in different steps of this process: selectins in rolling; chemokines (usually displayed bound to proteoglycans) in activating the neutrophils to increase avidity of integrins; integrins in firm adhesion; and CD31 (PECAM-1) in transmigration. ICAM-1, intercellular adhesion molecule-1; IL-1, interleukin-1; PECAM-1, platelet endothelial cell adhesion molecule-1; TNF, tumor necrosis factor.

Table 2–2 Endothelial and Leukocyte Adhesion Molecules

Endothelial Molecule	Leukocyte Molecule	Major Role(s)
Selectins and Selectin Ligands		
P-selectin	Sialyl–Lewis X–modified proteins	Rolling
E-selectin	Sialyl–Lewis X–modified proteins	Rolling and adhesion
GlyCam-1, CD34	L-selectin*	Rolling (neutrophils, monocytes)
Integrins and Integrin Ligands		
ICAM-1 (immunoglobulin family)	CD11/CD18 integrins (LFA-1, Mac-1)	Firm adhesion, arrest, transmigration
VCAM-1 (immunoglobulin family)	VLA-4 integrin	Adhesion
Others		
CD31	CD31 (homotypic interaction)	Transmigration of leukocytes through endothelium

*L-selectin is also involved in the binding of circulating lymphocytes to the high endothelial venules in lymph nodes and mucosal lymphoid tissues, and subsequent homing of lymphocytes to these tissues.
ICAM-1, intercellular adhesion molecule-1; LFA-1, leukocyte function–associated antigen-1; Mac-1, macrophage-1 antigen; VCAM-1, vascular cell adhesion molecule-1; VLA-4, very late antigen-4.

transmembrane heterodimeric glycoproteins that mediate the adhesion of leukocytes to endothelium and of various cells to the extracellular matrix. They are normally expressed on leukocyte plasma membranes in a low-affinity form and do not adhere to their specific ligands until the leukocytes are activated by chemokines.

Chemokines are chemoattractant cytokines that are secreted by many cells at sites of inflammation and are displayed on the endothelial surface. (Cytokines are described later in the chapter.) When the adherent leukocytes encounter the displayed chemokines, the cells are activated, and their integrins undergo conformational changes and cluster together, thus converting to a high-affinity form. At the same time, other cytokines, notably TNF and IL-1 (also secreted at sites of infection and injury), activate endothelial cells to increase their expression of ligands for integrins. These ligands include intercellular adhesion molecule-1 (ICAM-1), which binds to the integrins leukocyte function–associated antigen-1 (LFA-1) (also called CD11a/CD18) and macrophage-1 antigen (Mac-1) (i.e., CD11b/CD18), and vascular cell adhesion molecule-1 (VCAM-1), which binds to the integrin very late antigen-4 (VLA-4) (Table 2–2). Engagement of integrins by their ligands delivers signals to the leukocytes that lead to cytoskeletal changes that mediate firm attachment to the substrate. Thus, the net result of cytokine-stimulated increased integrin affinity and increased expression of integrin ligands is stable attachment of leukocytes to endothelial cells at sites of inflammation.

Transmigration. After being arrested on the endothelial surface, leukocytes migrate through the vessel wall primarily by squeezing between cells at intercellular junctions. This extravasation of leukocytes, called *diapedesis,* occurs mainly in the venules of the systemic vasculature; it has also been noted in capillaries in the pulmonary circulation. Migration of leukocytes is driven by chemokines produced in extravascular tissues, which stimulate movement of the leukocytes toward their chemical gradient. In addition, platelet endothelial cell adhesion molecule-1 (PECAM-1) (also called CD31), a cellular adhesion molecule expressed on leukocytes and endothelial cells, mediates the binding events needed for leukocytes to traverse the endothelium. After passing through the endothelium, leukocytes secrete collagenases that enable them to pass through the vascular basement membrane.

Chemotaxis. After extravasating from the blood, leukocytes move toward sites of infection or injury along a chemical gradient by a process called *chemotaxis*. Both exogenous and endogenous substances can be chemotactic for leukocytes, including the following:

- Bacterial products, particularly peptides with *N*-formylmethionine termini
- Cytokines, especially those of the *chemokine* family
- Components of the complement system, particularly C5
- Products of the lipoxygenase pathway of arachidonic acid (AA) metabolism, particularly leukotriene B4 (LTB4)

These mediators, which are described in more detail later, are produced in response to infections and tissue damage and during immunologic reactions. Leukocyte infiltration in all of these situations results from the actions of various combinations of mediators.

Chemotactic molecules bind to specific cell surface receptors, which triggers the assembly of cytoskeletal contractile elements necessary for movement. Leukocytes move by extending pseudopods that anchor to the ECM and then pull the cell in the direction of the extension. The direction of such movement is specified by a higher density of chemokine receptors at the leading edge of the cell. Thus, leukocytes move to and are retained at the site where they are needed.

The type of emigrating leukocyte varies with the age of the inflammatory response and with the type of stimulus. In most forms of acute inflammation, *neutrophils predominate in the inflammatory infiltrate during the first 6 to 24 hours and are replaced by monocytes in 24 to 48 hours* (Fig. 2–6). Several factors account for this early abundance of neutrophils: These cells are the most numerous leukocytes in the blood, they respond more rapidly to chemokines, and they may attach more firmly to the adhesion molecules that are rapidly induced on endothelial cells, such as P- and E-selectins. In addition, after entering tissues, neutrophils are short-lived—they die by apoptosis and disappear within 24 to 48 hours—while monocytes survive longer. There are exceptions to this pattern of cellular infiltration, however.

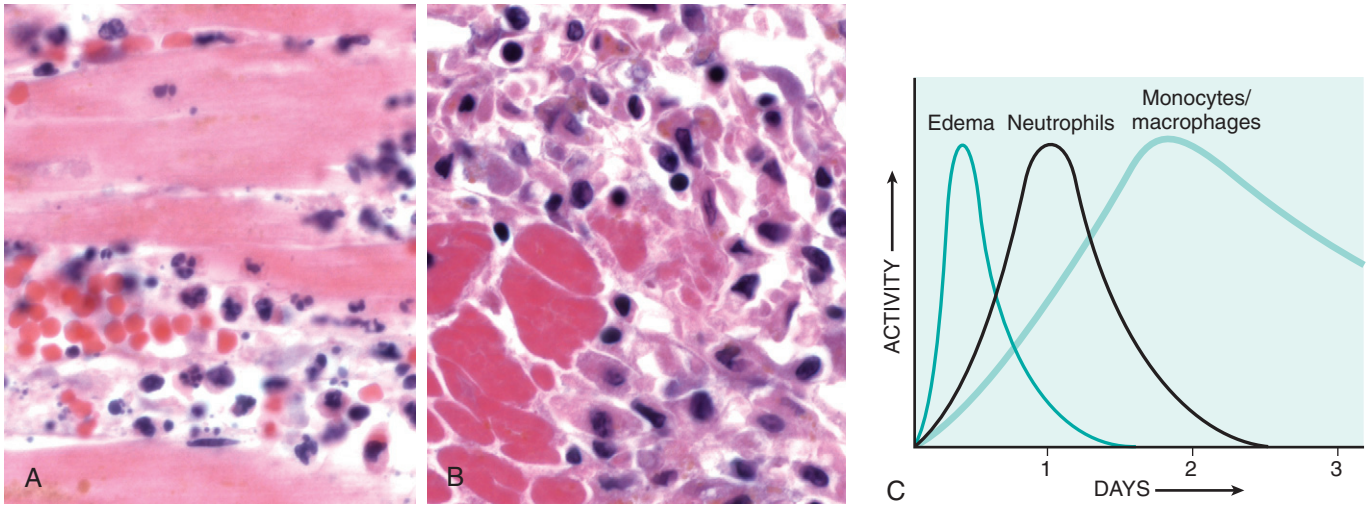


Figure 2–6 Nature of leukocyte infiltrates in inflammatory reactions. The photomicrographs show an inflammatory reaction in the myocardium after ischemic necrosis (infarction). **A,** Early (neutrophilic) infiltrates and congested blood vessels. **B,** Later (mononuclear) cellular infiltrates. **C,** The approximate kinetics of edema and cellular infiltration. For sake of simplicity, edema is shown as an acute transient response, although secondary waves of delayed edema and neutrophil infiltration also can occur.

In certain infections (e.g., those caused by *Pseudomonas* organisms), the cellular infiltrate is dominated by continuously recruited neutrophils for several days; in viral infections, lymphocytes may be the first cells to arrive; and in some hypersensitivity reactions, eosinophils may be the main cell type.

SUMMARY

Leukocyte Recruitment to Sites of Inflammation

- Leukocytes are recruited from the blood into the extravascular tissue, where infectious pathogens or damaged tissues may be located, and are activated to perform their functions.
- Leukocyte recruitment is a multi-step process consisting of loose attachment to and rolling on endothelium (mediated by selectins); firm attachment to endothelium (mediated by integrins); and migration through interendothelial spaces.
- Various cytokines promote expression of selectins and integrin ligands on endothelium (TNF, IL-1), increase the avidity of integrins for their ligands (chemokines), and promote directional migration of leukocytes (also chemokines); many of these cytokines are produced by tissue macrophages and other cells responding to pathogens or damaged tissues.
- Neutrophils predominate in the early inflammatory infiltrate and are later replaced by macrophages.

Leukocyte Activation

Once leukocytes have been recruited to the site of infection or tissue necrosis, they must be activated to perform their functions. Stimuli for activation include microbes, products of necrotic cells, and several mediators that are described later. As described earlier, leukocytes use various receptors to sense the presence of microbes, dead cells, and foreign substances. Engagement of these cellular receptors induces a number of responses in leukocytes that are part of their normal defensive functions and are grouped under the term *leukocyte activation* (Fig. 2–7). Leukocyte activation results in the enhancement of the following functions:

- *Phagocytosis* of particles
- *Intracellular destruction of phagocytosed microbes and dead cells* by substances produced in phagosomes, including reactive oxygen and nitrogen species and lysosomal enzymes
- *Liberation of substances that destroy extracellular microbes and dead tissues*, which are largely the same as the substances produced within phagocytic vesicles. A recently discovered mechanism by which neutrophils destroy extracellular microbes is the formation of extracellular "traps."
- *Production of mediators*, including arachidonic acid metabolites and cytokines, that amplify the inflammatory reaction, by recruiting and activating more leukocytes

Phagocytosis. *Phagocytosis consists of three steps* (Fig. 2–8): *(1) recognition and attachment of the particle to the ingesting leukocyte; (2) engulfment, with subsequent formation of a phagocytic vacuole; and (3) killing and degradation of the ingested material.*

Leukocytes bind and ingest most microorganisms and dead cells by means of specific surface receptors. Some of these receptors recognize components of the microbes and dead cells and other receptors recognize host proteins, called *opsonins*, that coat microbes and target them for phagocytosis (the process called *opsonization*). The most important opsonins are antibodies of the immunoglobulin G (IgG) class that bind to microbial surface antigens, breakdown products of the complement protein C3 (described later), and plasma carbohydrate-binding lectins called collectins, which bind to microbial cell wall sugar groups. These opsonins either are present in the blood ready to coat

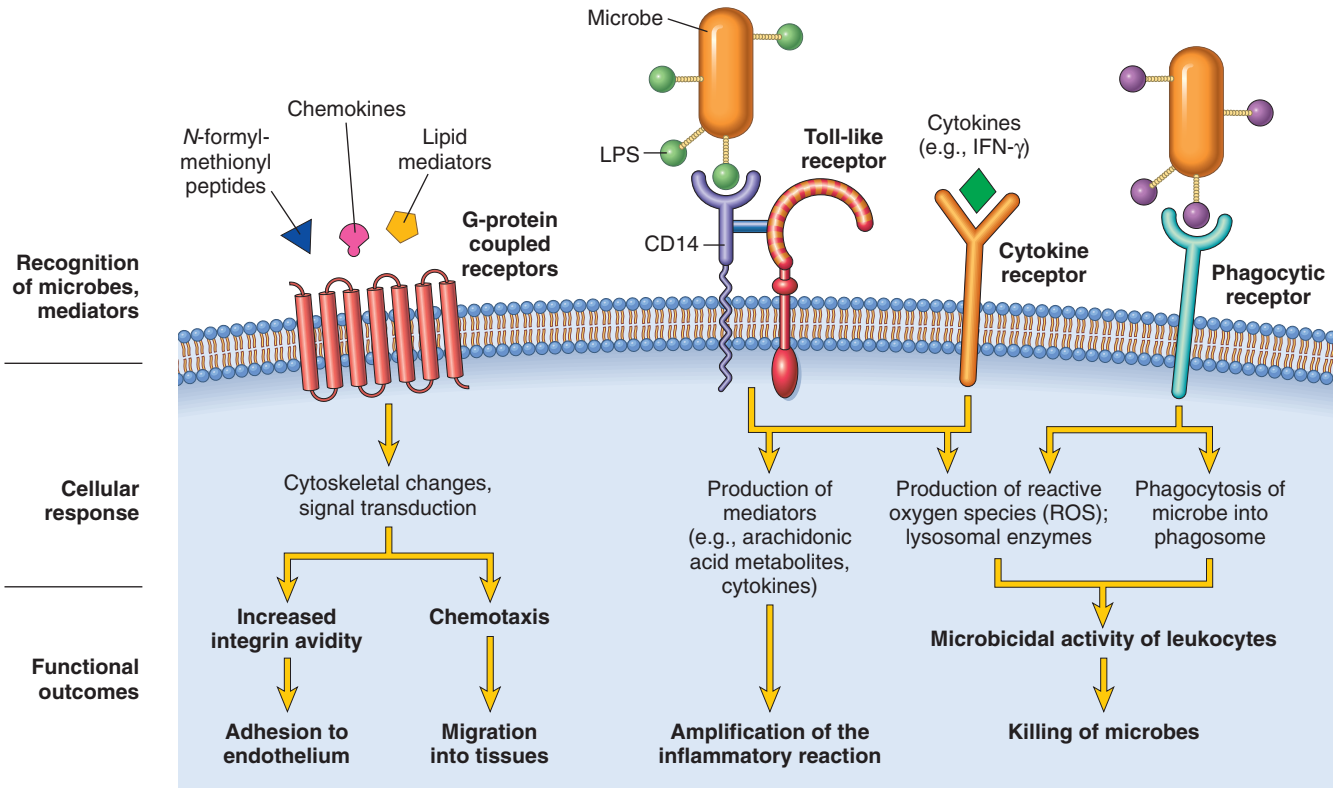


Figure 2–7 Leukocyte activation. Different classes of cell surface receptors of leukocytes recognize different stimuli. The receptors initiate responses that mediate the functions of the leukocytes. Only some receptors are depicted (see text for details). Lipopolysaccharide (LPS) first binds to a circulating LPS-binding protein (*not shown*). IFN-γ, interferon-γ.

microbes or are produced in response to the microbes. Leukocytes express receptors for opsonins that facilitate rapid phagocytosis of the coated microbes. These receptors include the Fc receptor for IgG (called FcγRI), complement receptors 1 and 3 (CR1 and CR3) for complement fragments, and C1q for the collectins.

Binding of opsonized particles to these receptors triggers engulfment and induces cellular activation that enhances degradation of ingested microbes. In engulfment, pseudopods are extended around the object, eventually forming a phagocytic vacuole. The membrane of the vacuole then fuses with the membrane of a lysosomal granule, resulting in discharge of the granule's contents into the *phagolysosome*.

Killing and Degradation of Phagocytosed Microbes. The culmination of the phagocytosis of microbes is killing and degradation of the ingested particles. The key steps in this reaction are the production of microbicidal substances within lysosomes and fusion of the lysosomes with phagosomes, thus exposing the ingested particles to the destructive mechanisms of the leukocytes (Fig. 2–8). The most important microbicidal substances are reactive oxygen species (ROS) and lysosomal enzymes. The production of ROS involves the following steps:

- Phagocytosis and the engagement of various cellular receptors stimulate an *oxidative burst,* also called the *respiratory burst,* which is characterized by a rapid increase in oxygen consumption, glycogen catabolism (glycogenolysis), increased glucose oxidation, and production of ROS. The generation of the oxygen metabolites is due to rapid activation of a leukocyte NADPH oxidase, called the *phagocyte oxidase,* which oxidizes NADPH (reduced nicotinamide adenine dinucleotide phosphate) and, in the process, converts oxygen to superoxide ion ($O_2^{\cdot-}$) (see Fig. 1–18, *B*, Chapter 1).
- Superoxide is then converted by spontaneous dismutation into hydrogen peroxide ($O_2^{\cdot-} + 2H^+ \rightarrow H_2O_2$). These ROS act as free radicals and destroy microbes by mechanisms that were described in Chapter 1.
- The quantities of H_2O_2 produced generally are insufficient to kill most bacteria (although superoxide and hydroxyl radical formation may be sufficient to do so). However, the lysosomes of neutrophils (called *azurophilic granules*) contain the enzyme myeloperoxidase (MPO), and in the presence of a halide such as Cl^-, MPO converts H_2O_2 to $HOCl^{\cdot}$ (hypochlorous radical). $HOCl^{\cdot}$ is a powerful oxidant and antimicrobial agent (NaOCl is the active ingredient in chlorine bleach) that kills bacteria by halogenation, or by protein and lipid peroxidation.

Fortunately, the phagocyte oxidase is active only after its cytosolic subunit translocates to the membrane of the phagolysosome; thus, the reactive end products are generated mainly within the vesicles, and the phagocyte itself is not damaged. H_2O_2 is eventually broken down to water and O_2 by the actions of catalase, and the other ROS also are degraded (Chapter 1). Reactive nitrogen species, particularly nitric oxide (NO), act in the same way as that described for ROS.

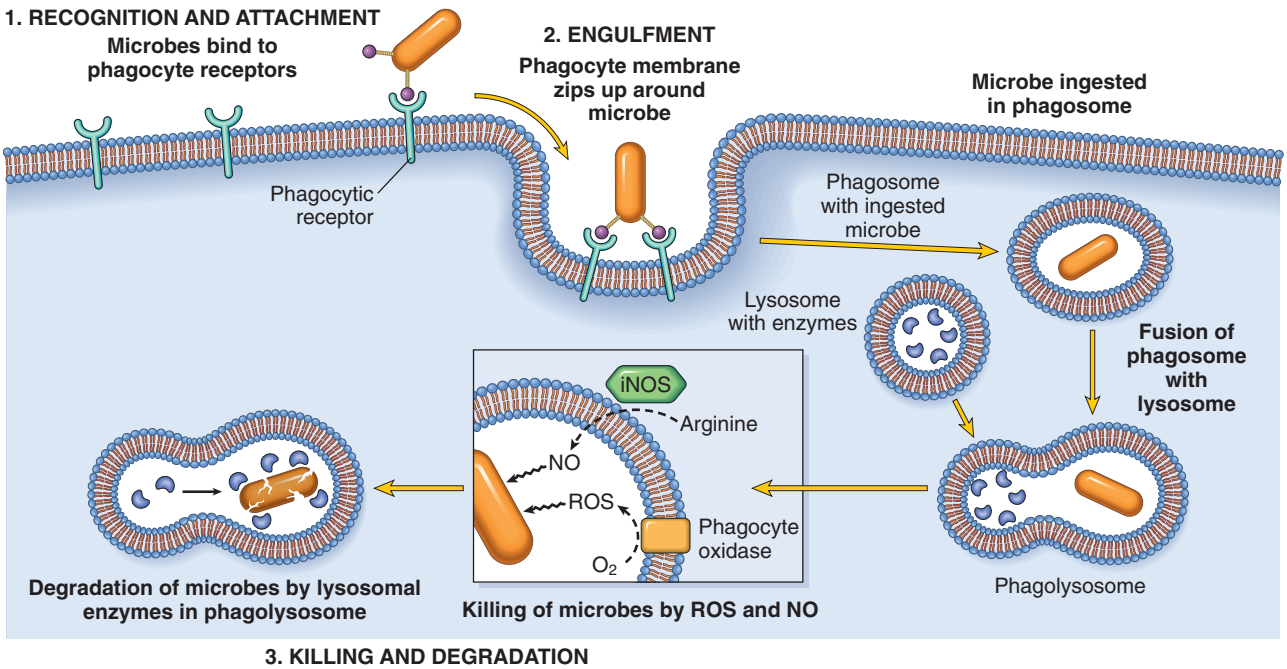


Figure 2–8 Phagocytosis. Phagocytosis of a particle (e.g., a bacterium) involves (*1*) attachment and binding of the particle to receptors on the leukocyte surface, (*2*) engulfment and fusion of the phagocytic vacuole with granules (lysosomes), and (*3*) destruction of the ingested particle. iNOS, inducible nitric oxide synthase; NO, nitric oxide; ROS, reactive oxygen species.

The dead microorganisms are then degraded by the action of lysosomal acid hydrolases. Perhaps the most important lysosomal enzyme involved in bacterial killing is elastase.

Of note, in addition to ROS and enzymes, several other constituents of leukocyte granules are capable of killing infectious pathogens. These include bactericidal permeability-increasing protein (causing phospholipase activation and membrane phospholipid degradation), lysozyme (causing degradation of bacterial coat oligosaccharides), major basic protein (an important eosinophil granule constituent that is cytotoxic for parasites), and defensins (peptides that kill microbes by creating holes in their membranes).

Secretion of Microbicidal Substances. The microbicidal mechanisms of phagocytes are largely sequestered within phagolysosomes in order to protect the leukocytes from damaging themselves. Leukocytes also actively secrete granule components including enzymes such as elastase, which destroy and digest extracellular microbes and dead tissues, as well as antimicrobial peptides. The contents of lysosomal granules are secreted by leukocytes into the extracellular milieu by several mechanisms:

- The phagocytic vacuole may remain transiently open to the outside before complete closure of the phagolysosome (regurgitation during feeding).
- If cells encounter materials that cannot be easily ingested, such as immune complexes deposited on immovable surfaces (e.g., glomerular basement membrane), the attempt to phagocytose these substances (frustrated phagocytosis) triggers strong leukocyte activation, and lysosomal enzymes are released into the surrounding tissue or lumen.
- The membrane of the phagolysosome may be damaged if potentially injurious substances, such as silica particles, are phagocytosed.

Neutrophil Extracellular Traps (NETs). These "traps" are extracellular fibrillar networks that are produced by neutrophils in response to infectious pathogens (mainly bacteria and fungi) and inflammatory mediators (such as chemokines, cytokines, complement proteins, and ROS). NETs contain a framework of nuclear chromatin with embedded granule proteins, such as antimicrobial peptides and enzymes (Fig. 2–9). The traps provide a high concentration of antimicrobial substances at sites of infection, and prevent the spread of the microbes by trapping them in the fibrils. In the process, the nuclei of the neutrophils are lost, leading to death of the cells. NETs also have been detected in blood neutrophils during sepsis. The nuclear chromatin in the NETs, which includes histones and associated DNA, has been postulated to be a source of nuclear antigens in systemic autoimmune diseases, particularly lupus, in which affected persons react against their own DNA and nucleoproteins (Chapter 4).

Leukocyte-Induced Tissue Injury

Because leukocytes are capable of secreting potentially harmful substances such as enzymes and ROS, they are important causes of injury to normal cells and tissues under several circumstances:

- As part of a normal defense reaction against infectious microbes, when "bystander" tissues are injured. In certain infections that are difficult to eradicate, such as tuberculosis and some viral diseases, the host response

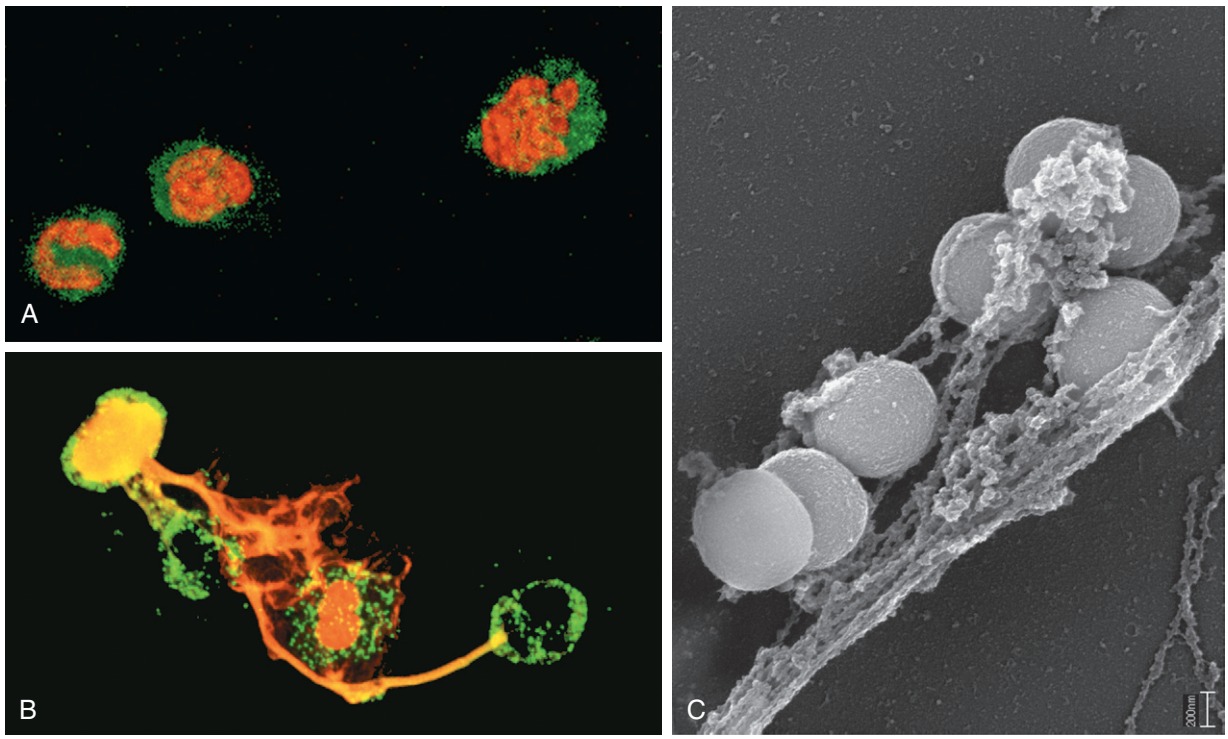


Figure 2–9 Neutrophil extracellular traps (NETs). **A,** Healthy neutrophils with nuclei stained red and cytoplasm green. **B,** Release of nuclear material from neutrophils (note that two have lost their nuclei), forming extracellular traps. **C,** An electron micrograph of bacteria (staphylococci) trapped in NETs.

(From Brinkmann V, Zychlinsky A: Beneficial suicide: why neutrophils die to make NETs. Nat Rev Microbiol 5:577, 2007, with the permission of the authors and publisher.)

contributes more to the pathologic process than does the microbe itself.

- As a normal attempt to clear damaged and dead tissues (e.g., after a myocardial infarction). In an infarct, inflammation may prolong and exacerbate the injurious consequences of the ischemia, especially upon reperfusion (Chapter 1).
- When the inflammatory response is inappropriately directed against host tissues, as in certain autoimmune diseases, or when the host reacts excessively against nontoxic environmental substances, such as allergic diseases including asthma (discussed in Chapter 4)

In all of these situations, the mechanisms by which leukocytes damage normal tissues are the same as the mechanisms involved in the clearance of microbes and dead tissues, because once the leukocytes are activated, their effector mechanisms do not distinguish between offender and host. In fact, if unchecked or inappropriately directed against host tissues, leukocytes themselves become the main offenders. Leukocyte-dependent tissue injury underlies many acute and chronic human diseases (Table 2–3), as is evident in discussions of specific disorders throughout this book.

Activated leukocytes, especially macrophages, also secrete many cytokines, which stimulate further inflammation and have important systemic effects, to be discussed later.

SUMMARY

Leukocyte Effector Mechanisms

- Leukocytes can eliminate microbes and dead cells by phagocytosis, followed by their destruction in phagolysosomes.
- Destruction is caused by free radicals (ROS, NO) generated in activated leukocytes and lysosomal enzymes.
- Enzymes and ROS may be released into the extracellular environment.
- The mechanisms that function to eliminate microbes and dead cells (the physiologic role of inflammation) are also capable of damaging normal tissues (the pathologic consequences of inflammation).

Defects in Leukocyte Function

Since leukocytes play a central role in host defense, it is not surprising that defects in leukocyte function, both acquired and inherited, lead to increased susceptibility to infections, which may be recurrent and life-threatening (Table 2–4). The most common causes of defective inflammation are bone marrow suppression caused by tumors or treatment with chemotherapy or radiation (resulting in decreased leukocyte numbers) and metabolic diseases such as

Table 2–3 Clinical Examples of Leukocyte-Induced Injury

Disorder*	Cells and Molecules Involved in Injury
Acute	
Acute respiratory distress syndrome	Neutrophils
Acute transplant rejection	Lymphocytes; antibodies and complement
Asthma	Eosinophils; IgE antibodies
Glomerulonephritis	Antibodies and complement; neutrophils, monocytes
Septic shock	Cytokines
Chronic	
Rheumatoid arthritis	Lymphocytes, macrophages; antibodies?
Asthma	Eosinophils; IgE antibodies
Atherosclerosis	Macrophages; lymphocytes?
Chronic transplant rejection	Lymphocytes, macrophages; cytokines
Pulmonary fibrosis	Macrophages; fibroblasts

*Listed are selected examples of diseases in which the host response plays a significant role in tissue injury. Some, such as asthma, can manifest with acute inflammation or a chronic illness with repeated bouts of acute exacerbation. These diseases and their pathogenesis are discussed in much more detail in relevant chapters.
IgE, immunoglobulin E.

Table 2–4 Defects in Leukocyte Functions

Disease	Defect
Acquired	
Bone marrow suppression: tumors (including leukemias), radiation, and chemotherapy	Production of leukocytes
Diabetes, malignancy, sepsis, chronic dialysis	Adhesion and chemotaxis
Anemia, sepsis, diabetes, malnutrition	Phagocytosis and microbicidal activity
Genetic	
Leukocyte adhesion deficiency 1	Defective leukocyte adhesion because of mutations in β chain of CD11/CD18 integrins
Leukocyte adhesion deficiency 2	Defective leukocyte adhesion because of mutations in fucosyl transferase required for synthesis of sialylated oligosaccharide (receptor for selectins)
Chronic granulomatous disease	Decreased oxidative burst
X-linked	Phagocyte oxidase (membrane component)
Autosomal recessive	Phagocyte oxidase (cytoplasmic components)
Myeloperoxidase deficiency	Decreased microbial killing because of defective MPO–H_2O_2 system
Chédiak-Higashi syndrome	Decreased leukocyte functions because of mutations affecting protein involved in lysosomal membrane traffic

H_2O_2, hydrogen peroxide; MPO, myeloperoxidase.
Modified from Gallin JI: Disorders of phagocytic cells. In Gallin JI, et al (eds): Inflammation: Basic Principles and Clinical Correlates, 2nd ed. New York, Raven Press, 1992, pp 860, 861.

diabetes (causing abnormal leukocyte functions). These are described elsewhere in the book.

The genetic disorders, although individually rare, illustrate the importance of particular molecular pathways in the complex inflammatory response. Some of the better understood inherited diseases are the following:

- *Defects in leukocyte adhesion.* In *leukocyte adhesion deficiency type 1 (LAD-1)*, defective synthesis of the CD18 β subunit of the leukocyte integrins LFA-1 and Mac-1 leads to impaired leukocyte adhesion to and migration through endothelium, and defective phagocytosis and generation of an oxidative burst. *Leukocyte adhesion deficiency type 2 (LAD-2)* is caused by a defect in fucose metabolism resulting in the absence of sialyl–Lewis X, the oligosaccharide on leukocytes that binds to selectins on activated endothelium. Its clinical manifestations are similar to but milder than those of LAD-1.
- *Defects in microbicidal activity.* An example is *chronic granulomatous disease,* a genetic deficiency in one of the several components of the phagocyte oxidase enzyme that is responsible for generating ROS. In these patients, engulfment of bacteria does not result in activation of oxygen-dependent killing mechanisms. In an attempt to control these infections, the microbes are surrounded by activated macrophages, forming the "granulomas" (see later) that give the disease its distinctive pathologic features and its somewhat misleading name.
- *Defects in phagolysosome formation.* One such disorder, *Chédiak-Higashi syndrome,* is an autosomal recessive disease that results from disordered intracellular trafficking of organelles, ultimately impairing the fusion of lysosomes with phagosomes. The secretion of lytic secretory granules by cytotoxic T lymphocytes is also affected, explaining the severe immunodeficiency typical of the disorder.
- Rare patients with defective host defenses have been shown to carry *mutations in TLR signaling pathways*. Inherited defects in components of adaptive immune responses also result in increased susceptibility to infections. These are described in Chapter 4.
- *Gain-of-function mutations in genes encoding some components of the inflammasome*, one of which is called *cryopyrin,* are responsible for rare but serious diseases called cryopyrin-associated periodic fever syndromes (CAPSs), which manifest with unrelenting fevers and other signs of inflammation and respond well to treatment with IL-1 antagonists.

Outcomes of Acute Inflammation

Although the consequences of acute inflammation are modified by the nature and intensity of the injury, the site and tissue affected, and the ability of the host to mount a response, *acute inflammation generally has one of three outcomes* (Fig. 2–10):

- *Resolution: Regeneration and repair.* When the injury is limited or short-lived, when there has been no or minimal tissue damage, and when the injured tissue is capable of regenerating, the usual outcome is restoration

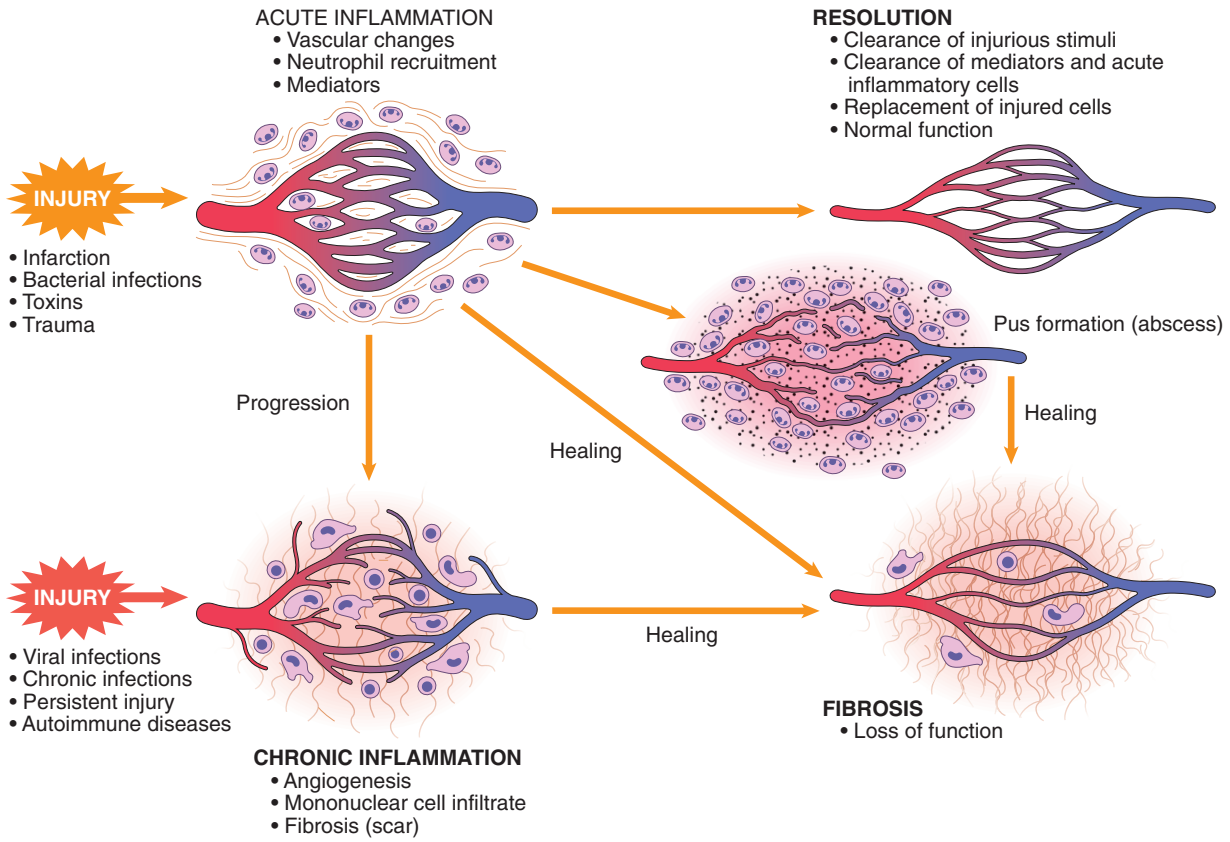


Figure 2–10 Outcomes of acute inflammation: resolution, healing by scarring (fibrosis), or chronic inflammation (see text).

to structural and functional normalcy. Before the process of resolution can start, the acute inflammatory response has to be terminated. This involves neutralization, decay, or enzymatic degradation of the various chemical mediators; normalization of vascular permeability; and cessation of leukocyte emigration, with subsequent death (by apoptosis) of extravasated neutrophils. Furthermore, leukocytes begin to produce mediators that inhibit inflammation, thereby limiting the reaction. The necrotic debris, edema fluid, and inflammatory cells are cleared by phagocytes and lymphatic drainage, eliminating the detritus from the battlefield. Leukocytes secrete cytokines that initiate the subsequent repair process, in which new blood vessels grow into the injured tissue to provide nutrients, growth factors stimulate the proliferation of fibroblasts and laying down of collagen to fill defects, and residual tissue cells proliferate to restore structural integrity. This process is described later in the chapter.

- *Chronic inflammation* may follow acute inflammation if the offending agent is not removed, or it may be present from the onset of injury (e.g., in viral infections or immune responses to self-antigens). Depending on the extent of the initial and continuing tissue injury, as well as the capacity of the affected tissues to regrow, chronic inflammation may be followed by restoration of normal structure and function or may lead to scarring.
- *Scarring* is a type of repair after substantial tissue destruction (as in abscess formation, discussed later) or when inflammation occurs in tissues that do not regenerate, in which the injured tissue is filled in by connective tissue. In organs in which extensive connective tissue deposition occurs in attempts to heal the damage or as a consequence of chronic inflammation, the outcome is *fibrosis,* a process that can significantly compromise function.

SUMMARY

Sequence of Events in Acute Inflammation

- The vascular changes in acute inflammation are characterized by increased blood flow secondary to arteriolar and capillary bed dilation (erythema and warmth).
- Increased vascular permeability, as a consequence of either widening of interendothelial cell junctions of the venules or direct endothelial cell injury, results in an exudate of protein-rich extravascular fluid (tissue edema).
- The leukocytes, initially predominantly neutrophils, adhere to the endothelium via adhesion molecules and then leave the microvasculature and migrate to the site of injury under the influence of chemotactic agents.
- Phagocytosis, killing, and degradation of the offending agent follow.
- Genetic or acquired defects in leukocyte functions give rise to recurrent infections.
- The outcome of acute inflammation may be removal of the exudate with restoration of normal tissue architecture (resolution); transition to chronic inflammation; or extensive destruction of the tissue resulting in scarring.

MORPHOLOGIC PATTERNS OF ACUTE INFLAMMATION

The vascular and cellular reactions that characterize acute inflammation are reflected in the morphologic appearance of the reaction. The severity of the inflammatory response, its specific cause, and the particular tissue involved all can modify the basic morphology of acute inflammation, producing distinctive appearances. The importance of recognizing these morphologic patterns is that they are often associated with different etiology and clinical situations.

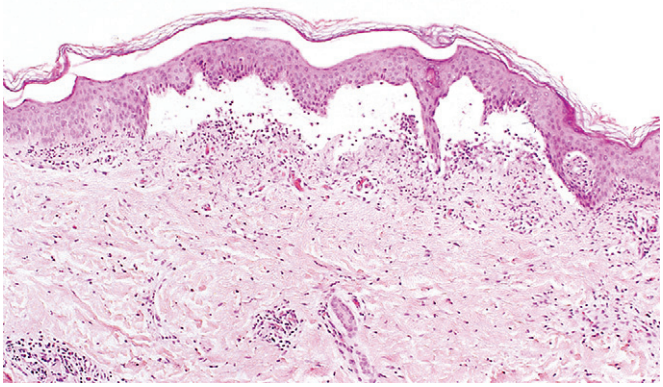


Figure 2–11 Serous inflammation. Low-power view of a cross-section of a skin blister showing the epidermis separated from the dermis by a focal collection of serous effusion.

MORPHOLOGY

- **Serous inflammation** is characterized by the outpouring of a watery, relatively protein-poor fluid that, depending on the site of injury, derives either from the plasma or from the secretions of mesothelial cells lining the peritoneal, pleural, and pericardial cavities. The skin blister resulting from a burn or viral infection is a good example of the accumulation of a serous effusion either within or immediately beneath the epidermis of the skin (Fig. 2–11). Fluid in a serous cavity is called an **effusion.**
- **Fibrinous inflammation** occurs as a consequence of more severe injuries, resulting in greater vascular permeability that allows large molecules (such as fibrinogen) to pass the endothelial barrier. Histologically, the accumulated extravascular fibrin appears as an eosinophilic meshwork of threads or sometimes as an amorphous coagulum (Fig. 2–12). A fibrinous exudate is characteristic of inflammation in the lining of body cavities, such as the meninges, pericardium, and pleura. Such exudates may be degraded by fibrinolysis, and the accumulated debris may be removed by macrophages, resulting in restoration of the normal tissue structure **(resolution).** However, extensive fibrin-rich exudates may not be completely removed, and are replaced by an ingrowth of fibroblasts and blood vessels **(organization),** leading ultimately to scarring that may have significant clinical consequences. For example, organization of a fibrinous pericardial exudate forms dense fibrous scar tissue that bridges or obliterates the pericardial space and restricts myocardial function.
- **Suppurative (purulent) inflammation and abscess formation.** These are manifested by the collection of large amounts of purulent exudate (pus) consisting of neutrophils, necrotic cells, and edema fluid. Certain organisms (e.g., staphylococci) are more likely to induce such localized suppuration and are therefore referred to as pyogenic (pus-forming). **Abscesses** are focal collections of pus that may be caused by seeding of pyogenic organisms into a tissue or by secondary infections of necrotic foci. Abscesses typically have a central, largely necrotic region rimmed by a layer of preserved neutrophils (Fig. 2–13), with a surrounding zone of dilated vessels and fibroblast proliferation indicative of attempted repair. As time passes, the abscess may become completely walled off and eventually be replaced by connective tissue. Because of the underlying tissue destruction, the usual outcome with abscess formation is scarring.

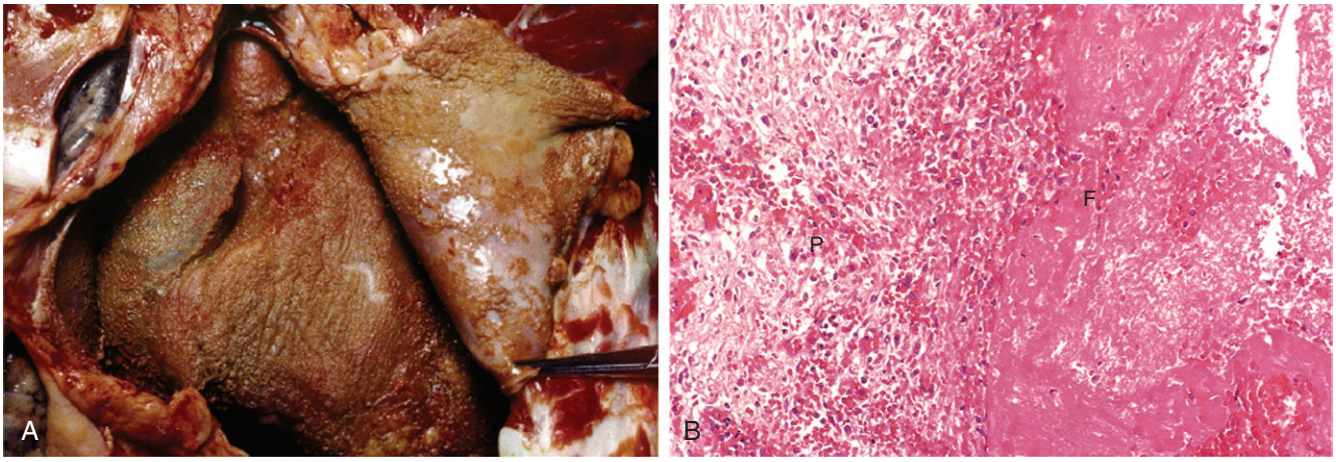


Figure 2–12 Fibrinous pericarditis. **A,** Deposits of fibrin on the pericardium. **B,** A pink meshwork of fibrin exudate *(F)* overlies the pericardial surface *(P)*.

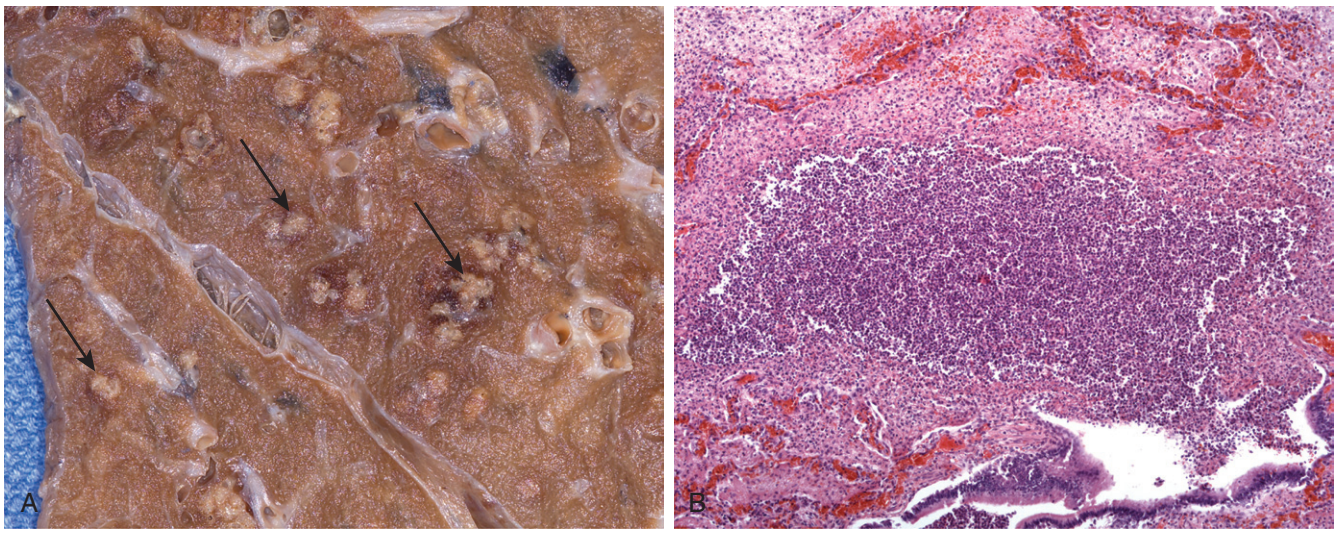


Figure 2–13 Purulent inflammation with abscess formation. **A,** Multiple bacterial abscesses in the lung (*arrows*) in a case of bronchopneumonia. **B,** The abscess contains neutrophils and cellular debris and is surrounded by congested blood vessels.

- An **ulcer** is a local defect, or excavation, of the surface of an organ or tissue that is produced by necrosis of cells and sloughing (shedding) of necrotic and inflammatory tissue (Fig. 2–14). Ulceration can occur only when tissue necrosis and resultant inflammation exist on or near a surface. Ulcers are most commonly encountered in (1) the mucosa of the mouth, stomach, intestines, or genitourinary tract and (2) in the subcutaneous tissues of the lower extremities in older persons who have circulatory disturbances predisposing affected tissue to extensive necrosis. Ulcerations are best exemplified by peptic ulcer of the stomach or duodenum, in which acute and chronic inflammation coexist. During the acute stage, there is intense polymorphonuclear infiltration and vascular dilation in the margins of the defect. With chronicity, the margins and base of the ulcer develop scarring with accumulation of lymphocytes, macrophages, and plasma cells.

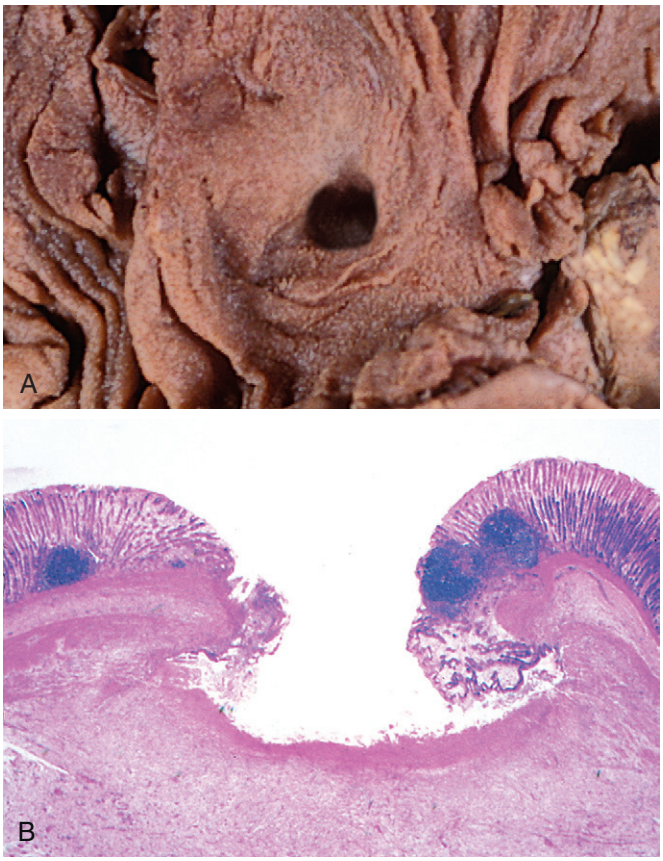


Figure 2–14 Ulcer. **A,** A chronic duodenal ulcer. **B,** Low-power cross-section of a duodenal ulcer crater with an acute inflammatory exudate in the base.

CHEMICAL MEDIATORS AND REGULATORS OF INFLAMMATION

Having described the vascular and cellular events in acute inflammation, and the accompanying morphologic alterations, we next discuss the chemical mediators that are responsible for these events. While the harried student may find this list daunting (as do the professors!), it is worthy of note that this knowledge has been used to design a large armamentarium of anti-inflammatory drugs, which are used every day by large numbers of people and include familiar drugs like aspirin and acetaminophen. In this section, we emphasize general properties of the mediators of inflammation and highlight only some of the more important molecules. We also touch upon some of the mechanisms that limit and terminate inflammatory reactions.

- *Mediators may be produced locally by cells at the site of inflammation, or may be derived from circulating inactive precursors (typically synthesized by the liver) that are activated at the site of inflammation* (Fig. 2–15 and Table 2–5). Cell-derived mediators are normally sequestered in intracellular granules and are rapidly secreted upon cellular activation (e.g., histamine in mast cells) or are

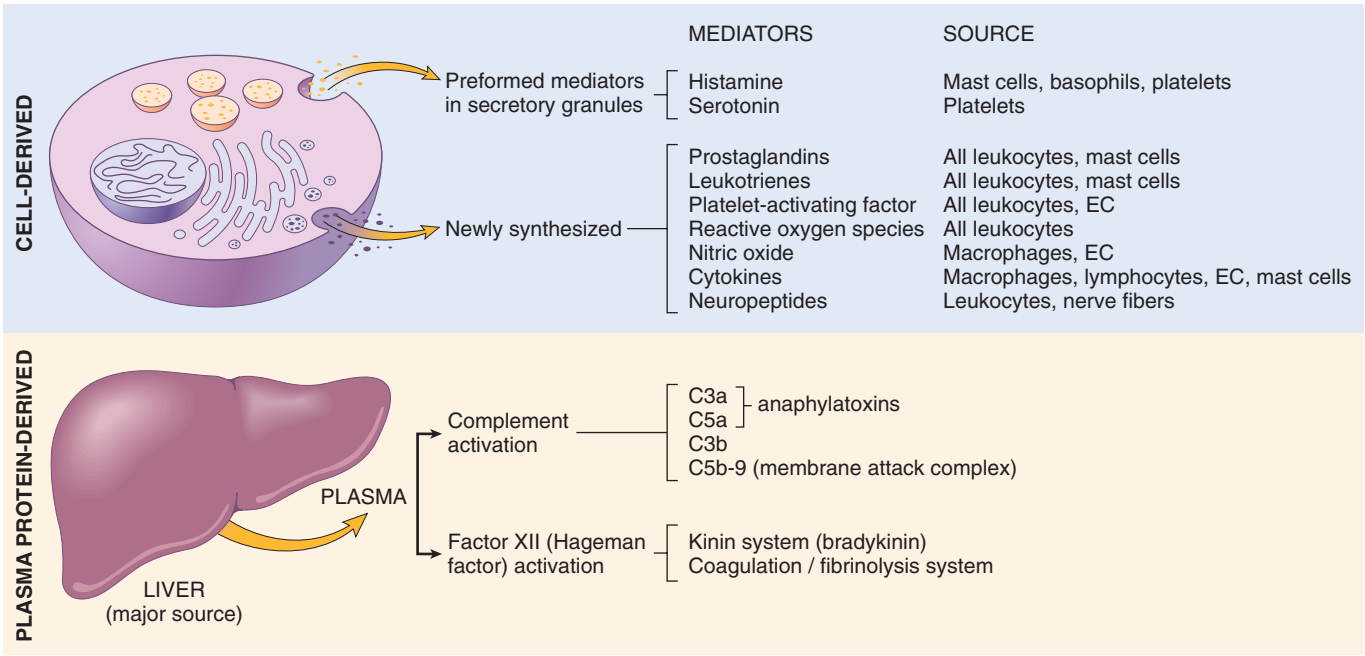


Figure 2–15 Mediators of inflammation. The principal cell-derived and plasma protein mediators are shown. EC, endothelial cells.

synthesized de novo in response to a stimulus (e.g., prostaglandins and cytokines produced by leukocytes and other cells). Plasma protein–derived mediators (complement proteins, kinins) circulate in an inactive form and typically undergo proteolytic cleavage to acquire their biologic activities.

- *Most mediators act by binding to specific receptors on different target cells.* Such mediators may act on only one or a very few cell types, or they may have diverse actions, with differing outcomes depending on which cell type they affect. Other mediators (e.g., lysosomal proteases, ROS) have direct enzymatic and/or toxic activities that do not require binding to specific receptors.
- *The actions of most mediators are tightly regulated and short-lived.* Once activated and released from the cell, mediators quickly decay (e.g., arachidonic acid metabolites), are inactivated by enzymes (e.g., kininase inactivates bradykinin), are eliminated (e.g., antioxidants scavenge

Table 2–5 Actions of the Principal Mediators of Inflammation

Mediator	Source(s)	Actions
Cell-Derived		
Histamine	Mast cells, basophils, platelets	Vasodilation, increased vascular permeability, endothelial activation
Serotonin	Platelets	Vasoconstriction
Prostaglandins	Mast cells, leukocytes	Vasodilation, pain, fever
Leukotrienes	Mast cells, leukocytes	Increased vascular permeability, chemotaxis, leukocyte adhesion and activation
Platelet-activating factor	Leukocytes, mast cells	Vasodilation, increased vascular permeability, leukocyte adhesion, chemotaxis, degranulation, oxidative burst
Reactive oxygen species	Leukocytes	Killing of microbes, tissue damage
Nitric oxide	Endothelium, macrophages	Vascular smooth muscle relaxation; killing of microbes
Cytokines (TNF, IL-1, IL-6)	Macrophages, endothelial cells, mast cells	*Local*: endothelial activation (expression of adhesion molecules). *Systemic*: fever, metabolic abnormalities, hypotension (shock)
Chemokines	Leukocytes, activated macrophages	Chemotaxis, leukocyte activation
Plasma Protein–Derived		
Complement	Plasma (produced in liver)	Leukocyte chemotaxis and activation, direct target killing (MAC), vasodilation (mast cell stimulation)
Kinins	Plasma (produced in liver)	Increased vascular permeability, smooth muscle contraction, vasodilation, pain
Proteases activated during coagulation	Plasma (produced in liver)	Endothelial activation, leukocyte recruitment

IL-1, IL-6, interleukin-1 and -6; MAC, membrane attack complex; TNF, tumor necrosis factor.

toxic oxygen metabolites), or are inhibited (e.g., complement regulatory proteins block complement activation).

Cell-Derived Mediators

Tissue macrophages, mast cells, and endothelial cells at the site of inflammation, as well as leukocytes that are recruited to the site from the blood, are all capable of producing different mediators of inflammation.

Vasoactive Amines

The two vasoactive amines, histamine and serotonin, are stored as preformed molecules in mast cells and other cells and are among the first mediators to be released in acute inflammatory reactions.

- *Histamine* is produced by many cell types, particularly mast cells adjacent to vessels, as well as circulating basophils and platelets. Preformed histamine is released from mast cell granules in response to a variety of stimuli: (1) physical injury such as trauma or heat; (2) immune reactions involving binding of IgE antibodies to Fc receptors on mast cells (Chapter 4); (3) C3a and C5a fragments of complement, the so-called anaphylatoxins (see later); (4) leukocyte-derived histamine-releasing proteins; (5) neuropeptides (e.g., substance P); and (6) certain cytokines (e.g., IL-1, IL-8). In humans, histamine causes arteriolar dilation and rapidly increases vascular permeability by inducing venular endothelial contraction and formation of interendothelial gaps. Soon after its release, histamine is inactivated by histaminase.
- *Serotonin* (5-hydroxytryptamine) is a preformed vasoactive mediator found within platelet granules that is released during platelet aggregation (Chapter 3). It induces vasoconstriction during clotting. It is produced mainly in some neurons and enterochromaffin cells, and is a neurotransmitter and regulates intestinal motility.

Arachidonic Acid Metabolites: Prostaglandins, Leukotrienes, and Lipoxins

Products derived from the metabolism of AA affect a variety of biologic processes, including inflammation and hemostasis. AA metabolites, also called *eicosanoids* (because they are derived from 20-carbon fatty acids—Greek *eicosa*, "twenty"), can mediate virtually every step of inflammation (Table 2–6); their synthesis is increased at sites of inflammatory response, and agents that inhibit their synthesis also diminish inflammation. Leukocytes, mast cells, endothelial cells, and platelets are the major sources of AA metabolites in inflammation. These AA-derived mediators act locally at the site of generation and then decay spontaneously or are enzymatically destroyed.

Table 2–6 Principal Inflammatory Actions of Arachidonic Acid Metabolites (Eicosanoids)

Action	Eicosanoid
Vasodilation	Prostaglandins PGI_2 (prostacyclin), PGE_1, PGE_2, PGD_2
Vasoconstriction	Thromboxane A_2, leukotrienes C_4, D_4, E_4
Increased vascular permeability	Leukotrienes C_4, D_4, E_4
Chemotaxis, leukocyte adhesion	Leukotriene B_4, HETE

HETE, hydroxyeicosatetraenoic acid.

AA is a 20-carbon polyunsaturated fatty acid (with four double bonds) produced primarily from dietary linoleic acid and present in the body mainly in its esterified form as a component of cell membrane phospholipids. It is released from these phospholipids through the action of cellular phospholipases that have been activated by mechanical, chemical, or physical stimuli, or by inflammatory mediators such as C5a. AA metabolism proceeds along one of two major enzymatic pathways: Cyclooxygenase stimulates the synthesis of prostaglandins and thromboxanes, and lipoxygenase is responsible for production of leukotrienes and lipoxins (Fig. 2–16).

- *Prostaglandins and thromboxanes.* Products of the cyclooxygenase pathway include prostaglandin E_2 (PGE_2), PGD_2, $PGF_{2\alpha}$, PGI_2 (prostacyclin), and thromboxane A_2 (TXA_2), each derived by the action of a specific enzyme on an intermediate. Some of these enzymes have a restricted tissue distribution. For example, platelets contain the enzyme *thromboxane synthase,* and hence TXA_2, a potent platelet-aggregating agent and vasoconstrictor, is the major prostaglandin produced in these cells. Endothelial cells, on the other hand, lack thromboxane synthase but contain prostacyclin synthase, which is responsible for the formation of PGI_2, a vasodilator and a potent inhibitor of platelet aggregation. The opposing roles of TXA_2 and PGI_2 in hemostasis are discussed further in Chapter 3. PGD_2 is the major metabolite of the cyclooxygenase pathway in mast cells; along with PGE_2 and $PGF_{2\alpha}$ (which are more widely distributed), it causes vasodilation and potentiates edema formation. The prostaglandins also contribute to the pain and fever that accompany inflammation; PGE_2 augments pain sensitivity to a variety of other stimuli and interacts with cytokines to cause fever.
- *Leukotrienes.* Leukotrienes are produced by the action of 5-lipoxygenase, the major AA-metabolizing enzyme in neutrophils. The synthesis of leukotrienes involves multiple steps (Fig. 2–16). The first step generates leukotriene A_4 (LTA_4), which in turn gives rise to LTB_4 or LTC_4. LTB_4 is produced by neutrophils and some macrophages and is a potent chemotactic agent for neutrophils. LTC_4 and its subsequent metabolites, LTD_4 and LTE_4, are produced mainly in mast cells and cause bronchoconstriction and increased vascular permeability.
- *Lipoxins.* Once leukocytes enter tissues, they gradually change their major lipoxygenase-derived AA products from leukotrienes to anti-inflammatory mediators called lipoxins, which inhibit neutrophil chemotaxis and adhesion to endothelium and thus serve as endogenous antagonists of leukotrienes. Platelets that are activated and adherent to leukocytes also are important sources of lipoxins. Platelets alone cannot synthesize lipoxins A_4 and B_4 (LXA_4 and LXB_4), but they can form these mediators from an intermediate derived from adjacent neutrophils, by a transcellular biosynthetic pathway. By this mechanism, AA products can pass from one cell type to another.

Anti-inflammatory Drugs That Block Prostaglandin Production. The central role of eicosanoids in inflammatory

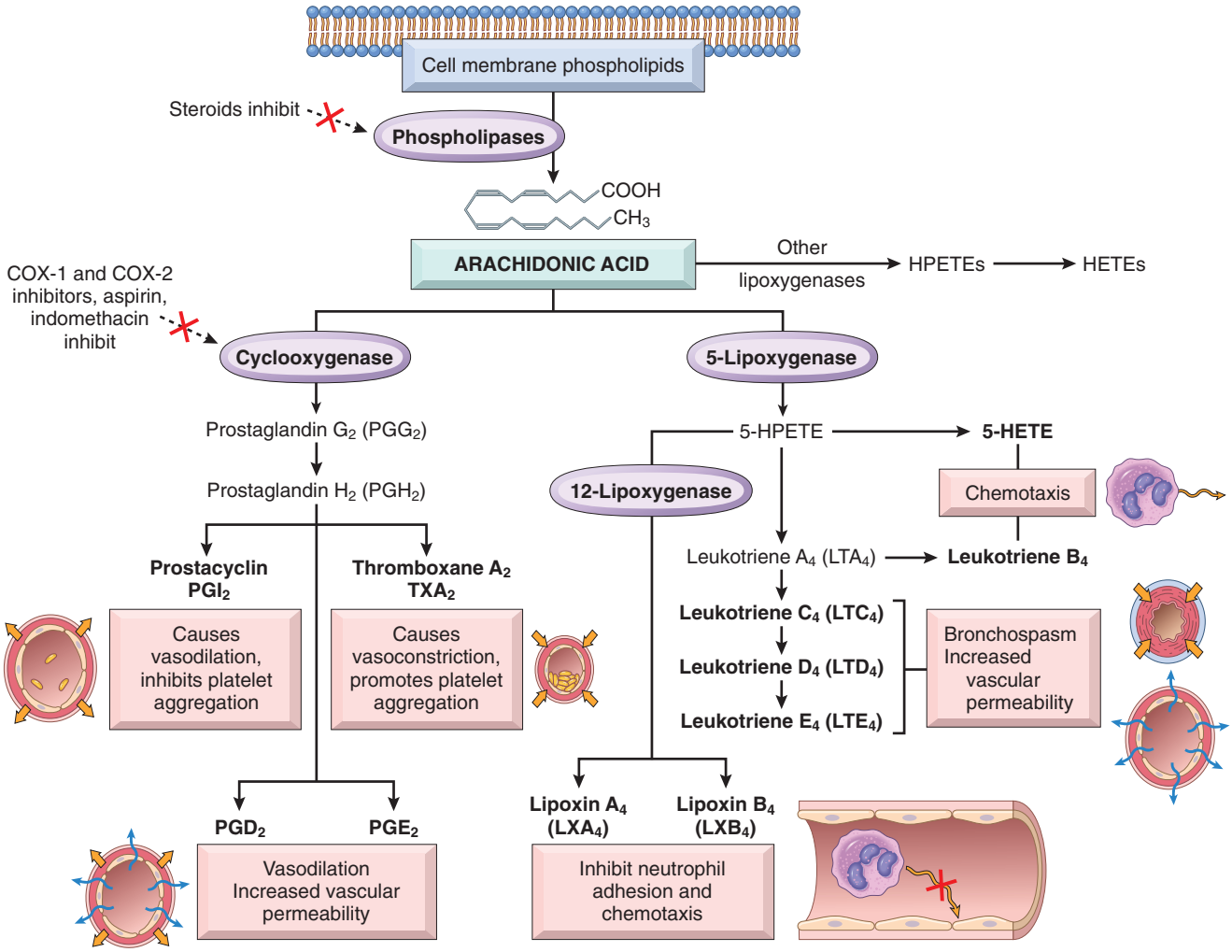


Figure 2–16 Production of arachidonic acid metabolites and their roles in inflammation. Note the enzymatic activities whose inhibition through pharmacologic intervention blocks major pathways (denoted with a *red X*). COX-1, COX-2, cyclooxygenases 1 and 2; HETE, hydroxyeicosatetraenoic acid; HPETE, hydroperoxyeicosatetraenoic acid.

processes is emphasized by the clinical utility of agents that block eicosanoid synthesis. Nonsteroidal anti-inflammatory drugs (NSAIDs), such as aspirin and ibuprofen, inhibit cyclooxygenase activity, thereby blocking all prostaglandin synthesis (hence their efficacy in treating pain and fever). There are two forms of the cyclooxygenase enzyme, COX-1 and COX-2. COX-1 is produced in response to inflammatory stimuli and also is constitutively expressed in most tissues, where it stimulates the production of prostaglandins that serve a homeostatic function (e.g., fluid and electrolyte balance in the kidneys, cytoprotection in the gastrointestinal tract). By contrast, COX-2 is induced by inflammatory stimuli but it is absent from most normal tissues. Therefore, COX-2 inhibitors have been developed with the expectation that they will inhibit harmful inflammation but will not block the protective effects of constitutively produced prostaglandins. These distinctions between the roles of the two cyclooxygenases are not absolute, however. Furthermore, COX-2 inhibitors may increase the risk for cardiovascular and cerebrovascular events, possibly because they impair endothelial cell production of prostacyclin (PGI_2), an inhibitor of platelet aggregation, but leave intact the COX-1–mediated production by platelets of TXA_2, a mediator of platelet aggregation. Glucocorticoids, which are powerful anti-inflammatory agents, act in part by inhibiting the activity of phospholipase A_2 and thus the release of AA from membrane lipids.

Platelet-Activating Factor

Originally named for its ability to aggregate platelets and cause their degranulation, platelet-activating factor (PAF) is another phospholipid-derived mediator with a broad spectrum of inflammatory effects. PAF is acetyl glycerol ether phosphocholine; it is generated from the membrane phospholipids of neutrophils, monocytes, basophils, endothelial cells, and platelets (and other cells) by the action of phospholipase A_2. PAF acts directly on target cells through the effects of a specific G protein–coupled receptor. In addition to stimulating platelets, PAF causes bronchoconstriction and is 100 to 1000 times more potent than histamine in inducing vasodilation and increased vascular permeability. It also stimulates the synthesis of other mediators, such as eicosanoids and cytokines, from platelets and other cells. Thus, PAF can elicit many of the reactions of inflammation,

including enhanced leukocyte adhesion, chemotaxis, leukocyte degranulation, and the respiratory burst.

Cytokines

Cytokines are polypeptide products of many cell types that function as mediators of inflammation and immune responses (Chapter 4). Different cytokines are involved in the earliest immune and inflammatory reactions to noxious stimuli and in the later adaptive (specific) immune responses to microbes. Some cytokines stimulate bone marrow precursors to produce more leukocytes, thus replacing the ones that are consumed during inflammation and immune responses. Molecularly characterized cytokines are called interleukins (abbreviated IL and numbered), referring to their ability to mediate communications between leukocytes. However, the nomenclature is imperfect—many interleukins act on cells other than leukocytes, and many cytokines that do act on leukocytes are not called interleukins, for historical reasons.

The major cytokines in acute inflammation are TNF, IL-1, IL-6, and a group of chemoattractant cytokines called chemokines. Other cytokines that are more important in chronic inflammation include interferon-γ (IFN-γ) and IL-12. A cytokine called IL-17, produced by T lymphocytes and other cells, plays an important role in recruiting neutrophils and is involved in host defense against infections and in inflammatory diseases.

Tumor Necrosis Factor and Interleukin-1. TNF and IL-1 are produced by activated macrophages, as well as mast cells, endothelial cells, and some other cell types (Fig. 2–17). Their secretion is stimulated by microbial products, such as bacterial endotoxin, immune complexes, and products of T lymphocytes generated during adaptive immune responses. As mentioned earlier, IL-1 is also the cytokine induced by activation of the inflammasome. The principal role of these cytokines in inflammation is in endothelial activation. Both TNF and IL-1 stimulate the expression of adhesion molecules on endothelial cells, resulting in increased leukocyte binding and recruitment, and enhance the production of additional cytokines (notably chemokines) and eicosanoids. TNF also increases the thrombogenicity of endothelium. IL-1 activates tissue fibroblasts, resulting in increased proliferation and production of ECM.

Although TNF and IL-1 are secreted by macrophages and other cells at sites of inflammation, they may enter the circulation and act at distant sites to induce the systemic acute-phase reaction that is often associated with infection and inflammatory diseases. Components of this reaction include fever, lethargy, hepatic synthesis of various acute-phase proteins (also stimulated by IL-6), metabolic wasting (cachexia), neutrophil release into the circulation, and fall in blood pressure. These systemic manifestations of inflammation are described later in the chapter.

Chemokines. The chemokines are a family of small (8 to 10 kDa), structurally related proteins that act primarily as chemoattractants for different subsets of leukocytes. The two main functions of chemokines are to recruit leukocytes to the site of inflammation and to control the normal anatomic organization of cells in lymphoid and other tissues. Combinations of chemokines that are produced transiently in response to inflammatory stimuli recruit particular cell

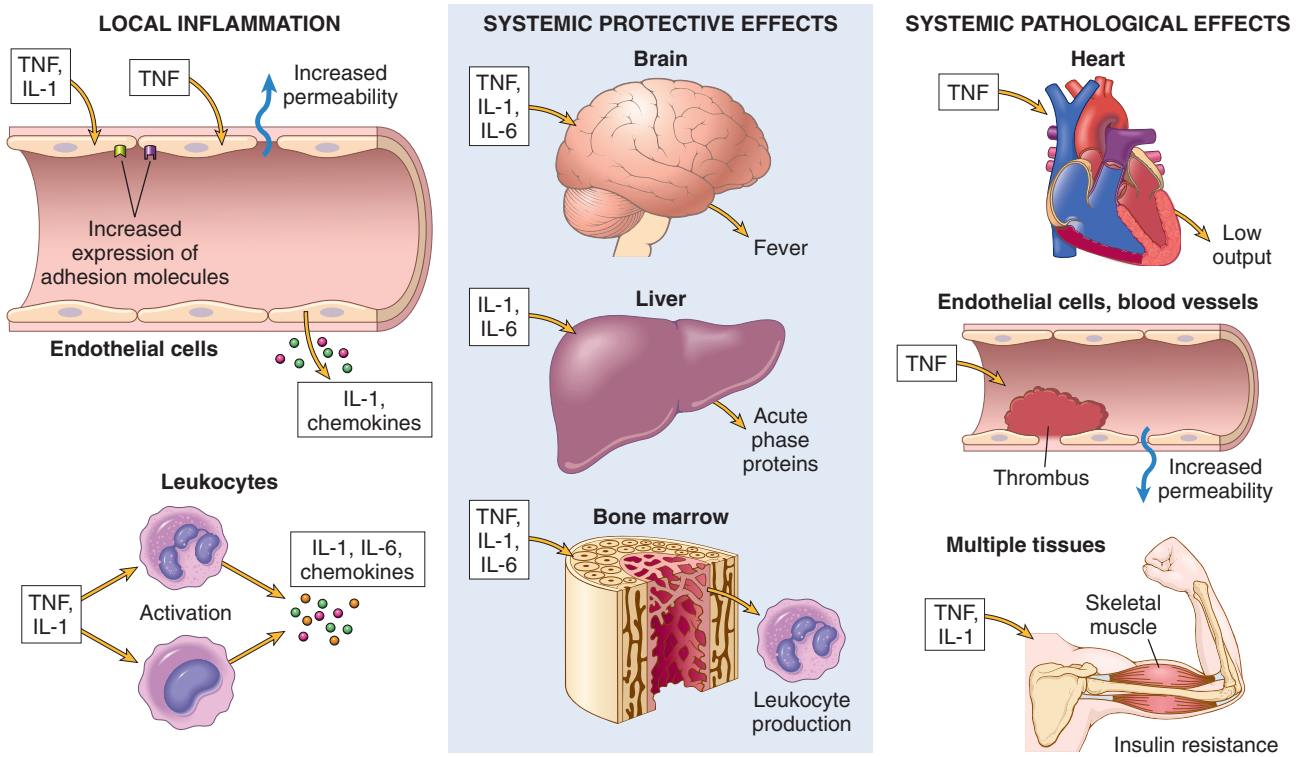


Figure 2–17 The roles of cytokines in acute inflammation. The cytokines TNF, IL-1, and IL-6 are key mediators of leukocyte recruitment in local inflammatory responses and also play important roles in the systemic reactions of inflammation.

populations (e.g., neutrophils, lymphocytes or eosinophils) to sites of inflammation. Chemokines also activate leukocytes; one consequence of such activation, as mentioned earlier, is increased affinity of leukocyte integrins for their ligands on endothelial cells. Some chemokines are produced constitutively in tissues and are responsible for the anatomic segregation of different cell populations in tissues (e.g., the segregation of T and B lymphocytes in different areas of lymph nodes and spleen). Chemokines mediate their activities by binding to specific G protein–coupled receptors on target cells; two of these chemokine receptors (called CXCR4 and CCR5) are important coreceptors for the binding and entry of the human immunodeficiency virus into lymphocytes (Chapter 4).

Chemokines are classified into four groups based on the arrangement of conserved cysteine residues. The two major groups are the CXC and CC chemokines:

- *CXC chemokines* have one amino acid separating the conserved cysteines and act primarily on neutrophils. IL-8 is typical of this group; it is produced by activated macrophages, endothelial cells, mast cells, and fibroblasts, mainly in response to microbial products and other cytokines such as IL-1 and TNF.
- *CC chemokines* have adjacent cysteine residues and include monocyte chemoattractant protein-1 (MCP-1) and macrophage inflammatory protein-1α (MIP-1α) (both chemotactic predominantly for monocytes), RANTES (*r*egulated on *a*ctivation, *n*ormal *T* cell–*e*xpressed and *s*ecreted) (chemotactic for memory CD4+ T cells and monocytes), and eotaxin (chemotactic for eosinophils).

Reactive Oxygen Species

ROS are synthesized via the NADPH oxidase (phagocyte oxidase) pathway and are released from neutrophils and macrophages that are activated by microbes, immune complexes, cytokines, and a variety of other inflammatory stimuli. The synthesis and regulation of these oxygen-derived free radicals have been described in Chapter 1, in the context of cell injury, and earlier in this chapter in the discussion of leukocyte activation. When the ROS are produced within lysosomes they function to destroy phagocytosed microbes and necrotic cells. When secreted at low levels, ROS can increase chemokine, cytokine, and adhesion molecule expression, thus amplifying the cascade of inflammatory mediators. At higher levels, these mediators are responsible for tissue injury by several mechanisms, including (1) endothelial damage, with thrombosis and increased permeability; (2) protease activation and antiprotease inactivation, with a net increase in breakdown of the ECM; and (3) direct injury to other cell types (e.g., tumor cells, red cells, parenchymal cells). Fortunately, various antioxidant protective mechanisms (e.g., mediated by catalase, superoxide dismutase, and glutathione) present in tissues and blood minimize the toxicity of the oxygen metabolites (Chapter 1).

Nitric Oxide

NO is a short-lived, soluble, free radical gas produced by many cell types and capable of mediating a variety of functions. In the central nervous system it regulates neurotransmitter release as well as blood flow. Macrophages use it as a cytotoxic agent for killing microbes and tumor cells. When produced by endothelial cells it relaxes vascular smooth muscle and causes vasodilation.

NO is synthesized de novo from L-arginine, molecular oxygen, and NADPH by the enzyme nitric oxide synthase (NOS). There are three isoforms of NOS, with different tissue distributions.

- Type I, neuronal NOS (nNOS), is constitutively expressed in neurons, and does not play a significant role in inflammation.
- Type II, inducible NOS (iNOS), is induced in macrophages and endothelial cells by a number of inflammatory cytokines and mediators, most notably by IL-1, TNF, and IFN-γ, and by bacterial endotoxin, and is responsible for production of NO in inflammatory reactions. This inducible form is also present in many other cell types, including hepatocytes, cardiac myocytes, and respiratory epithelial cells.
- Type III, endothelial NOS, (eNOS), is constitutively synthesized primarily (but not exclusively) in endothelium.

An important function of NO is as a microbicidal (cytotoxic) agent in activated macrophages. NO plays other roles in inflammation, including vasodilation, antagonism of all stages of platelet activation (adhesion, aggregation, and degranulation), and reduction of leukocyte recruitment at inflammatory sites.

Lysosomal Enzymes of Leukocytes

The lysosomal granules of neutrophils and monocytes contain many enzymes that destroy phagocytosed substances and are capable of causing tissue damage. Lysosomal granule contents also may be released from activated leukocytes, as described earlier. Acid proteases generally are active only in the low-pH environment of phagolysosomes, whereas neutral proteases, including elastase, collagenase, and cathepsin, are active in extracellular locations and cause tissue injury by degrading elastin, collagen, basement membrane, and other matrix proteins. Neutral proteases also can cleave the complement proteins C3 and C5 directly to generate the vasoactive mediators C3a and C5a and can generate bradykinin-like peptides from kininogen.

The potentially damaging effects of lysosomal enzymes are limited by antiproteases present in the plasma and tissue fluids. These include α_1-antitrypsin, the major inhibitor of neutrophil elastase, and α_2-macroglobulin. Deficiencies of these inhibitors may result in sustained activation of leukocyte proteases, resulting in tissue destruction at sites of leukocyte accumulation. For instance, α_1-antitrypsin deficiency in the lung can cause a severe panacinar emphysema (Chapter 12).

Neuropeptides

Like the vasoactive amines, neuropeptides can initiate inflammatory responses; these are small proteins, such as substance P, that transmit pain signals, regulate vessel tone, and modulate vascular permeability. Nerve fibers that secrete neuropeptides are especially prominent in the lung and gastrointestinal tract.

SUMMARY

Major Cell-Derived Mediators of Inflammation

- Vasoactive amines—histamine, serotonin: Their main effects are vasodilation and increased vascular permeability.
- Arachidonic acid metabolites—prostaglandins and leukotrienes: Several forms exist and are involved in vascular reactions, leukocyte chemotaxis, and other reactions of inflammation; they are antagonized by lipoxins.
- Cytokines: These proteins, produced by many cell types, usually act at short range; they mediate multiple effects, mainly in leukocyte recruitment and migration; principal ones in acute inflammation are TNF, IL-1, IL-6, and chemokines.
- ROS: Roles include microbial killing and tissue injury.
- NO: Effects are vasodilation and microbial killing.
- Lysosomal enzymes: Roles include microbial killing and tissue injury.

Plasma Protein–Derived Mediators

Circulating proteins of three interrelated systems—the complement, kinin, and coagulation systems—are involved in several aspects of the inflammatory reaction.

Complement

The *complement system* consists of plasma proteins that play an important role in host defense (immunity) and inflammation. Upon activation, different complement proteins coat (opsonize) particles, such as microbes, for phagocytosis and destruction, and contribute to the inflammatory response by increasing vascular permeability and leukocyte chemotaxis. Complement activation ultimately generates a porelike membrane attack complex (MAC) that punches holes in the membranes of invading microbes. Here we summarize the role of the complement system in inflammation.

- Complement components, numbered C1 to C9, are present in plasma in inactive forms, and many of them are activated by proteolysis to acquire their own proteolytic activity, thus setting up an enzymatic cascade.
- The critical step in the generation of biologically active complement products is the activation of the third component, C3 (Fig. 2–18). C3 cleavage occurs by three pathways: (1) the *classical pathway,* triggered by fixation of the first complement component C1 to antigen-antibody complexes; (2) the *alternative pathway,* triggered by bacterial polysaccharides (e.g., endotoxin) and other microbial cell wall components, and involving a distinct set of plasma proteins including properdin and factors B and D; and (3) the *lectin pathway,* in which a plasma lectin binds to mannose residues on microbes and activates an early component of the classical pathway (but in the absence of antibodies).
- All three pathways lead to the formation of a C3 convertase that cleaves C3 to C3a and C3b. C3b deposits on the cell or microbial surface where complement was activated and then binds to the C3 convertase complex to form C5 convertase; this complex cleaves C5 to generate C5a and C5b and initiate the final stages of assembly of C6 to C9.

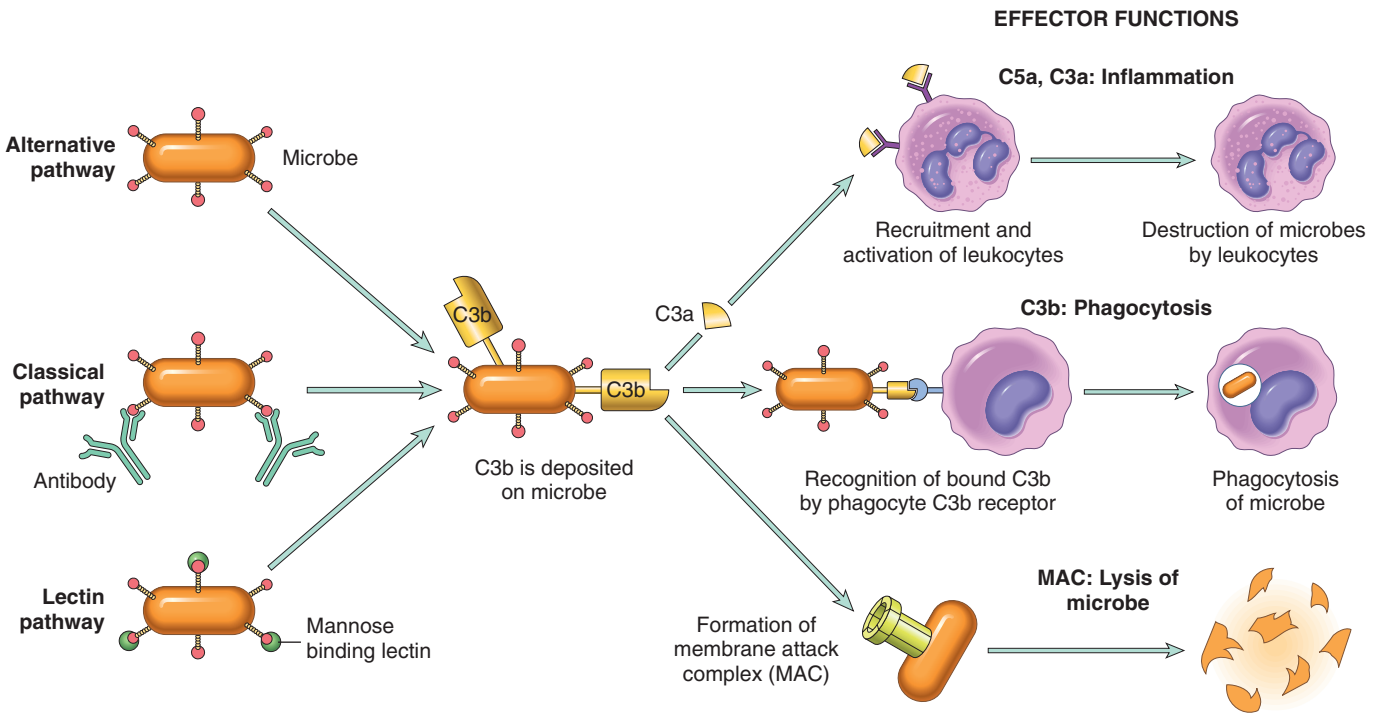


Figure 2–18 The activation and functions of the complement system. Activation of complement by different pathways leads to cleavage of C3. The functions of the complement system are mediated by breakdown products of C3 and other complement proteins, and by the membrane attack complex (MAC).

The complement-derived factors that are produced along the way contribute to a variety of phenomena in acute inflammation:

- *Vascular effects.* C3a and C5a increase vascular permeability and cause vasodilation by inducing mast cells to release histamine. These complement products are also called anaphylatoxins because their actions mimic those of mast cells, which are the main cellular effectors of the severe allergic reaction called anaphylaxis (Chapter 4). C5a also activates the lipoxygenase pathway of AA metabolism in neutrophils and macrophages, causing release of more inflammatory mediators.
- *Leukocyte activation, adhesion, and chemotaxis.* C5a, and to lesser extent, C3a and C4a, activate leukocytes, increasing their adhesion to endothelium, and is a potent chemotactic agent for neutrophils, monocytes, eosinophils, and basophils.
- *Phagocytosis.* When fixed to a microbial surface, C3b and its inactive proteolytic product iC3b act as opsonins, augmenting phagocytosis by neutrophils and macrophages, which express receptors for these complement products.
- The MAC, which is made up of multiple copies of the final component C9, kills some bacteria (especially thin-walled *Neisseria*) by creating pores that disrupt osmotic balance.

The activation of complement is tightly controlled by cell-associated and circulating regulatory proteins. The presence of these inhibitors in host cell membranes protects normal cells from inappropriate damage during protective reactions against microbes. Inherited deficiencies of these regulatory proteins lead to spontaneous complement activation:

- A protein called *C1 inhibitor* blocks activation of C1, and its inherited deficiency causes a disease called hereditary angioedema, in which excessive production of kinins secondary to complement activation results in edema in multiple tissues, including the larynx.
- Another protein called *decay-accelerating factor* (DAF) normally limits the formation of C3 and C5 convertases. In a disease called *paroxysmal nocturnal hemoglobinuria,* there is an acquired deficiency of DAF that results in complement-mediated lysis of red cells (which are more sensitive to lysis than most nucleated cells) (Chapter 11).
- *Factor H* is a plasma protein that also limits convertase formation; its deficiency is associated with a kidney disease called the *hemolytic uremic syndrome* (Chapter 13), as well as spontaneous vascular permeability in *macular degeneration* of the eye.

Even in the presence of regulatory proteins, inappropriate or excessive complement activation (e.g., in antibody-mediated diseases) can overwhelm the regulatory mechanisms; this is why complement activation is responsible for serious tissue injury in a variety of immunologic disorders (Chapter 4).

Coagulation and Kinin Systems

Some of the molecules activated during blood clotting are capable of triggering multiple aspects of the inflammatory response. *Hageman factor* (also known as *factor XII of the intrinsic coagulation cascade*) (Fig. 2–19) is a protein synthesized by the liver that circulates in an inactive form until it encounters collagen, basement membrane, or activated platelets (e.g., at a site of endothelial injury). Activated Hageman factor (factor XIIa) initiates four systems that may contribute to the inflammatory response: (1) the kinin system, producing vasoactive kinins; (2) the clotting

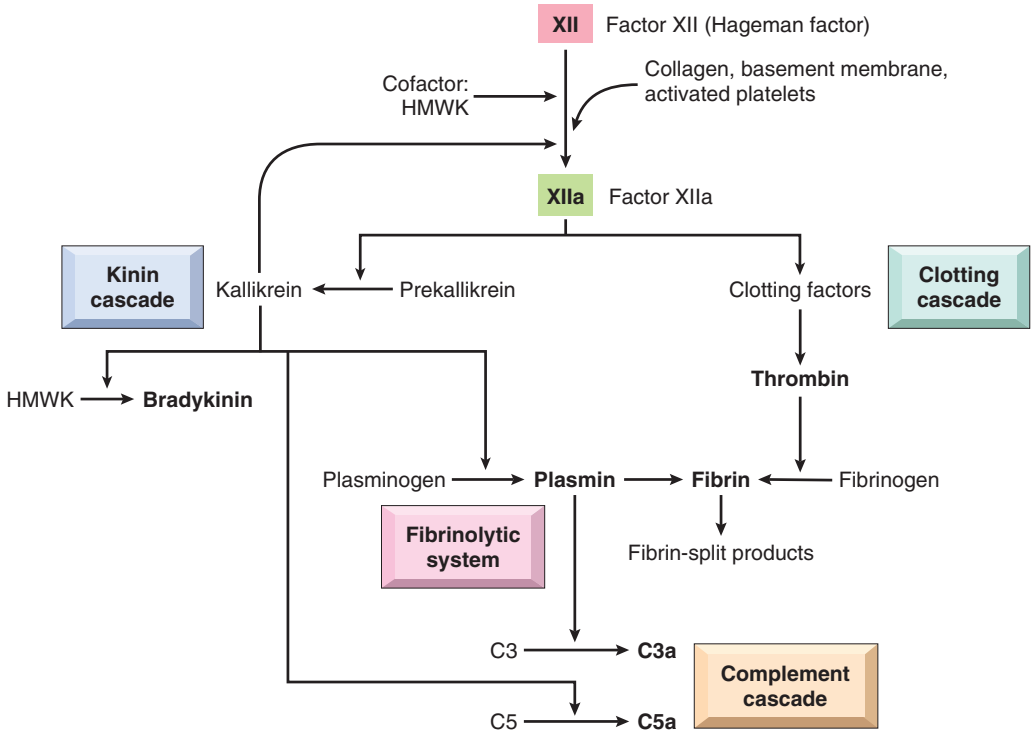


Figure 2–19 Interrelationships among the four plasma mediator systems triggered by activation of factor XII (Hageman factor). See text for details.

system, inducing the activation of thrombin, fibrinopeptides, and factor X, all with inflammatory properties; (3) the fibrinolytic system, producing plasmin and inactivating thrombin; and (4) the complement system, producing the anaphylatoxins C3a and C5a. These are described below.

- *Kinin system* activation leads ultimately to the formation of *bradykinin* from its circulating precursor, high-molecular-weight kininogen (HMWK) (Fig. 2–19). Like histamine, bradykinin causes increased vascular permeability, arteriolar dilation, and bronchial smooth muscle contraction. It also causes pain when injected into the skin. The actions of bradykinin are short-lived because it is rapidly degraded by kininases present in plasma and tissues. Of note, *kallikrein*, an intermediate in the kinin cascade with chemotactic activity, also is a potent activator of Hageman factor and thus constitutes another link between the kinin and clotting systems.
- In the *clotting system* (Chapter 3), the proteolytic cascade leads to activation of thrombin, which then cleaves circulating soluble fibrinogen to generate an insoluble *fibrin clot. Factor Xa,* an intermediate in the clotting cascade, causes increased vascular permeability and leukocyte emigration. Thrombin participates in inflammation by binding to protease-activated receptors that are expressed on platelets, endothelial cells, and many other cell types. Binding of thrombin to these receptors on endothelial cells leads to their activation and enhanced leukocyte adhesion. In addition, thrombin generates *fibrinopeptides* (during fibrinogen cleavage) that increase vascular permeability and are chemotactic for leukocytes. Thrombin also cleaves C5 to generate C5a, thus linking coagulation with complement activation.
- As a rule, whenever clotting is initiated (e.g., by activated Hageman factor), the *fibrinolytic system* is also activated concurrently. This mechanism serves to limit clotting by cleaving fibrin, thereby solubilizing the fibrin clot (Chapter 3). *Plasminogen activator* (released from endothelium, leukocytes, and other tissues) and *kallikrein* cleave *plasminogen,* a plasma protein bound up in the evolving fibrin clot. The resulting product, plasmin, is a multifunctional protease that cleaves fibrin and is therefore important in lysing clots. However, fibrinolysis also participates in multiple steps in the vascular phenomena of inflammation. For example, fibrin degradation products increase vascular permeability, and plasmin cleaves the C3 complement protein, resulting in production of C3a and vasodilation and increased vascular permeability. Plasmin can also activate Hageman factor, thereby amplifying the entire set of responses.

As is evident from the preceding discussion, many molecules are involved in different aspects of the inflammatory reaction, and these molecules often interact with, amplify, and antagonize one another. From this almost bewildering potpourri of chemical mediators, it is possible to identify the major contributors to various components of acute inflammation (Table 2–7). The relative contributions of individual mediators to inflammatory reactions to different stimuli have yet to be fully elucidated. Such knowledge would have obvious therapeutic implications since it might allow one to "custom design" antagonists for various inflammatory diseases.

Table 2–7 Role of Mediators in Different Reactions of Inflammation

Inflammatory Component	Mediators
Vasodilation	Prostaglandins Nitric oxide Histamine
Increased vascular permeability	Histamine and serotonin C3a and C5a (by liberating vasoactive amines from mast cells, other cells) Bradykinin Leukotrienes C_4, D_4, E_4 PAF Substance P
Chemotaxis, leukocyte recruitment and activation	TNF, IL-1 Chemokines C3a, C5a Leukotriene B_4 Bacterial products (e.g., *N*-formyl methyl peptides)
Fever	IL-1, TNF Prostaglandins
Pain	Prostaglandins Bradykinin
Tissue damage	Lysosomal enzymes of leukocytes Reactive oxygen species Nitric oxide

IL-1, interleukin-1; PAF, platelet-activating factor; TNF, tumor necrosis factor.

SUMMARY

Plasma Protein–Derived Mediators of Inflammation

- Complement proteins: Activation of the complement system by microbes or antibodies leads to the generation of multiple breakdown products, which are responsible for leukocyte chemotaxis, opsonization and phagocytosis of microbes and other particles, and cell killing.
- Coagulation proteins: Activated factor XII triggers the clotting, kinin, and complement cascades and activates the fibrinolytic system.
- Kinins: Produced by proteolytic cleavage of precursors, this group mediates vascular reaction and pain.

Anti-inflammatory Mechanisms

Inflammatory reactions subside because many of the mediators are short-lived and are destroyed by degradative *enzymes*. In addition, there are several mechanisms that counteract inflammatory mediators and function to limit or terminate the inflammatory response. Some of these, such as lipoxins, and complement regulatory proteins, have been mentioned earlier. Activated macrophages and other cells secrete a cytokine, IL-10, whose major function is to down-regulate the responses of activated macrophages, thus providing a negative feedback loop. In a rare inherited disease in which IL-10 receptors are mutated, affected patients develop severe colitis in infancy. Other anti-inflammatory cytokines include TGF-β, which is also a mediator of fibrosis in tissue repair after inflammation. Cells also express a number of intracellular proteins, such as tyrosine phosphatases, that inhibit pro-inflammatory

signals triggered by receptors that recognize microbes and cytokines.

CHRONIC INFLAMMATION

Chronic inflammation is inflammation of prolonged duration (weeks to years) in which continuing inflammation, tissue injury, and healing, often by fibrosis, proceed simultaneously. In contrast with acute inflammation, which is distinguished by vascular changes, edema, and a predominantly neutrophilic infiltrate, chronic inflammation is characterized by a different set of reactions (Fig. 2–20; see also Table 2–1):

- *Infiltration with mononuclear cells,* including macrophages, lymphocytes, and plasma cells
- *Tissue destruction,* largely induced by the products of the inflammatory cells
- *Repair,* involving new vessel proliferation (angiogenesis) and fibrosis

Acute inflammation may progress to chronic inflammation if the acute response cannot be resolved, either because of the persistence of the injurious agent or because of interference with the normal process of healing. For example, a peptic ulcer of the duodenum initially shows acute inflammation followed by the beginning stages of resolution. However, recurrent bouts of duodenal epithelial injury interrupt this process, resulting in a lesion characterized by both acute and chronic inflammation (Chapter 14). Alternatively, some forms of injury (e.g., immunologic reactions, some viral infections) engender a chronic inflammatory response from the outset.

Chronic inflammation may arise in the following settings:

- *Persistent infections* by microbes that are difficult to eradicate. These include *Mycobacterium tuberculosis, Treponema pallidum* (the causative organism of syphilis), and certain viruses and fungi, all of which tend to establish persistent infections and elicit a T lymphocyte–mediated immune response called *delayed-type hypersensitivity* (Chapter 4).
- *Immune-mediated inflammatory diseases (hypersensitivity diseases).* Diseases that are caused by excessive and inappropriate activation of the immune system are increasingly recognized as being important health problems (Chapter 4). Under certain conditions, immune reactions develop against the affected person's own tissues, leading to *autoimmune diseases.* In such diseases, autoantigens evoke a self-perpetuating immune reaction that results in tissue damage and persistent inflammation. Autoimmunity plays an important role in several common and debilitating chronic inflammatory diseases, such as rheumatoid arthritis, inflammatory bowel disease, and psoriasis. Immune responses against common environmental substances are the cause of *allergic diseases,* such as bronchial asthma. Immune-mediated diseases may show morphologic patterns of mixed acute and chronic inflammation because they are characterized by repeated bouts of inflammation. Since, in most cases, the eliciting antigens cannot be eliminated, these disorders tend to be chronic and intractable.
- *Prolonged exposure to potentially toxic agents.* Examples are nondegradable exogenous materials such as inhaled particulate silica, which can induce a chronic inflammatory response in the lungs (silicosis, Chapter 12), and endogenous agents such as cholesterol crystals, which may contribute to atherosclerosis (Chapter 9).
- Mild forms of chronic inflammation may be important in the pathogenesis of many diseases that are not conventionally thought of as inflammatory disorders. Such diseases include neurodegenerative disorders such as Alzheimer disease, atherosclerosis, metabolic syndrome and the associated type 2 diabetes, and some forms of cancer in which inflammatory reactions promote tumor development. As mentioned earlier in the chapter, in many of these conditions the inflammation may be triggered by recognition of the initiating stimuli by the inflammasome. The role of inflammation in these conditions is discussed in the relevant chapters.

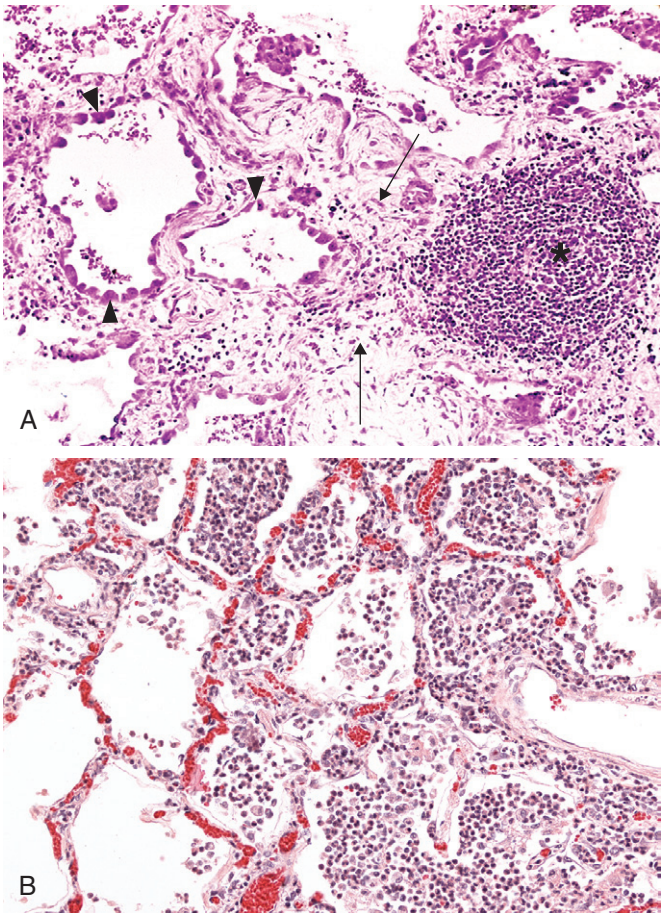


Figure 2–20 A, Chronic inflammation in the lung, showing the characteristic histologic features: collection of chronic inflammatory cells (*asterisk*); destruction of parenchyma, in which normal alveoli are replaced by spaces lined by cuboidal epithelium (*arrowheads*); and replacement by connective tissue, resulting in fibrosis (*arrows*). **B,** By contrast, in acute inflammation of the lung (acute bronchopneumonia), neutrophils fill the alveolar spaces and blood vessels are congested.

Chronic Inflammatory Cells and Mediators

The combination of prolonged and repeated inflammation, tissue destruction and fibrosis that characterizes chronic inflammation involves complex interactions between

several cell populations and their secreted mediators. Understanding the pathogenesis of chronic inflammatory reactions requires an appreciation of these cells and their biologic responses and functions.

Macrophages

Macrophages, the dominant cells of chronic inflammation, are tissue cells derived from circulating blood monocytes after their emigration from the bloodstream. Macrophages are normally diffusely scattered in most connective tissues and are also found in organs such as the liver (where they are called Kupffer cells), spleen and lymph nodes (where they are called sinus histiocytes), central nervous system (microglial cells), and lungs (alveolar macrophages). Together these cells constitute the so-called *mononuclear phagocyte system,* also known by the older name of reticuloendothelial system. In all tissues, macrophages act as filters for particulate matter, microbes, and senescent cells, as well as the effector cells that eliminate microbes in cellular and humoral immune responses (Chapter 4).

Monocytes arise from precursors in the bone marrow and circulate in the blood for only about a day. Under the influence of adhesion molecules and chemokines, they migrate to a site of injury within 24 to 48 hours after the onset of acute inflammation, as described earlier. When monocytes reach the extravascular tissue, they undergo transformation into macrophages, which are somewhat larger and have a longer lifespan and a greater capacity for phagocytosis than do blood monocytes.

Tissue macrophages are activated by diverse stimuli to perform a range of functions. Two major pathways of macrophage activation, *classical* and *alternative,* have been described (Fig. 2–21):

- *Classical macrophage activation* is induced by microbial products such as endotoxin, by T cell–derived signals, importantly the cytokine IFN-γ, and by foreign substances including crystals and particulate matter. Classically activated macrophages produce lysosomal enzymes, NO, and ROS, all of which enhance their ability to kill ingested organisms, and secrete cytokines that stimulate inflammation. These macrophages are important in host defense against ingested microbes and in many chronic inflammatory reactions.
- *Alternative macrophage activation* is induced by cytokines other than IFN-γ, such as IL-4 and IL-13, produced by T lymphocytes and other cells, including mast cells and eosinophils. Alternatively activated macrophages are not actively microbicidal; instead, their principal role is in tissue repair. They secrete growth factors that promote angiogenesis, activate fibroblasts and stimulate collagen synthesis. It may be that in response to most injurious stimuli, macrophages are initially activated by the classical pathway, designed to destroy the offending agents, and this is followed by alternative activation, which initiates tissue repair. However, such a precise sequence is not well documented in most inflammatory reactions.

Macrophages have several critical roles in host defense and the inflammatory response.

- Macrophages, like the other type of phagocyte, the neutrophils, *ingest and eliminate microbes and dead tissues.* Because macrophages respond to activating signals from T lymphocytes, they are the most important phagocytes in the cell-mediated arm of adaptive immune responses (Chapter 4).
- Macrophages *initiate the process of tissue repair* and are involved in scar formation and fibrosis.
- Macrophages *secrete mediators of inflammation*, such as cytokines (TNF, IL-1, chemokines, and others) and eicosanoids. These cells are therefore central to the initiation and propagation of all inflammatory reactions.
- Macrophages *display antigens to T lymphocytes and respond to signals from T cells,* thus setting up a feedback loop that

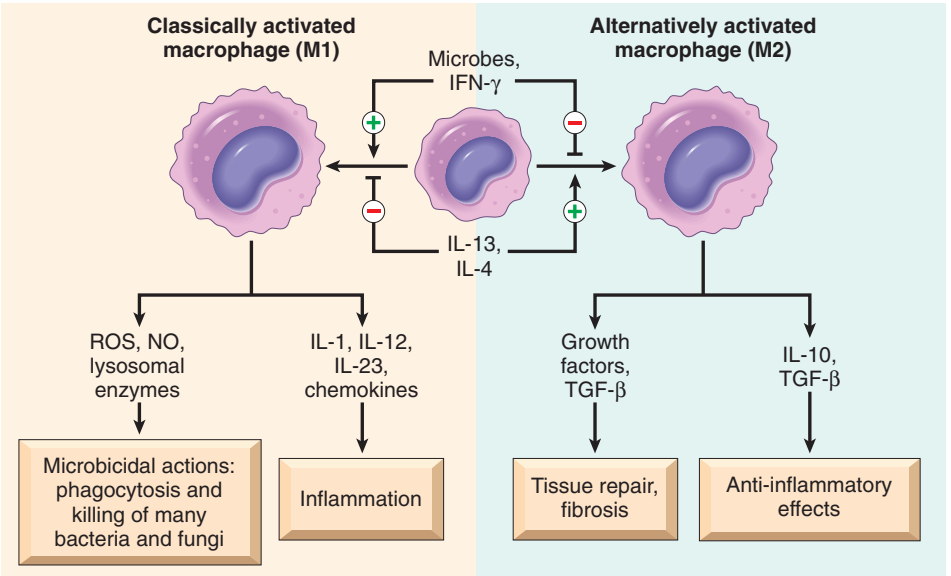


Figure 2–21 Pathways of macrophage activation. Different stimuli activate monocytes/macrophages to develop into functionally distinct populations. Classically activated macrophages are induced by microbial products and cytokines, particularly IFN-γ, and are microbicidal and involved in potentially harmful inflammation. Alternatively activated macrophages are induced by IL-4 and IL-13, produced by T_H2 cells (a helper T cell subset) and other leukocytes, and are important in tissue repair and fibrosis. IFN-γ, interferon-γ; IL-4, IL-13, interkeukin-4, -13.

is essential for defense against many microbes by cell-mediated immune responses. The same bidirectional interactions are central to the development of chronic inflammatory diseases. The roles of cytokines in these interactions are discussed later.

After the initiating stimulus is eliminated and the inflammatory reaction abates, macrophages eventually die or wander off into lymphatics. In chronic inflammatory sites, however, macrophage accumulation persists, because of continued recruitment from the blood and local proliferation. IFN-γ can also induce macrophages to fuse into large, multinucleate giant cells.

Lymphocytes

Lymphocytes are mobilized in the setting of any specific immune stimulus (i.e., infections) as well as non–immune-mediated inflammation (e.g., due to ischemic necrosis or trauma), and are the major drivers of inflammation in many autoimmune and other chronic inflammatory diseases. The activation of T and B lymphocytes is part of the adaptive immune response in infections and immunologic diseases (Chapter 4). Both classes of lymphocytes migrate into inflammatory sites using some of the same adhesion molecule pairs and chemokines that recruit other leukocytes. In the tissues, B lymphocytes may develop into *plasma cells,* which secrete antibodies, and CD4+ T lymphocytes are activated to secrete cytokines.

By virtue of cytokine secretion, CD4+ T lymphocytes promote inflammation and influence the nature of the inflammatory reaction. There are three subsets of CD4+ helper T cells that secrete different sets of cytokines and elicit different types of inflammation:

- T_H1 cells produce the cytokine IFN-γ, which activates macrophages in the classical pathway.
- T_H2 cells secrete IL-4, IL-5, and IL-13, which recruit and activate eosinophils and are responsible for the alternative pathway of macrophage activation.
- T_H17 cells secrete IL-17 and other cytokines that induce the secretion of chemokines responsible for recruiting neutrophils and monocytes into the reaction.

Both T_H1 and T_H17 cells are involved in defense against many types of bacteria and viruses and in autoimmune diseases. T_H2 cells are important in defense against helminthic parasites and in allergic inflammation. These T cell subsets and their functions are described in more detail in Chapter 4.

Lymphocytes and macrophages interact in a bidirectional way, and these interactions play an important role in propagating chronic inflammation (Fig. 2–22). Macrophages display antigens to T cells, express membrane molecules (called costimulators), and produce cytokines (IL-12 and others) that stimulate T cell responses (Chapter 4). Activated T lymphocytes, in turn, produce cytokines, described earlier, which recruit and activate macrophages and thus promote more antigen presentation and cytokine secretion. The result is a cycle of cellular reactions that fuel and sustain chronic inflammation. In some strong and prolonged inflammatory reactions, the accumulation of lymphocytes, antigen-presenting cells, and plasma cells may assume the morphologic features of lymphoid organs, and akin to lymph nodes, may even contain well-formed germinal centers. This pattern of lymphoid organogenesis is often seen in the synovium of patients with long-standing rheumatoid arthritis and the thyroid of patients with autoimmune thyroiditis.

Other Cells

Eosinophils are characteristically found in inflammatory sites around parasitic infections and as part of immune reactions mediated by IgE, typically associated with allergies. Their recruitment is driven by adhesion molecules similar to those used by neutrophils, and by specific chemokines (e.g., eotaxin) derived from leukocytes and epithelial cells. Eosinophil granules contain major basic protein, a highly charged cationic protein that is toxic to parasites but also causes epithelial cell necrosis.

Mast cells are sentinel cells widely distributed in connective tissues throughout the body, and they can participate in both acute and chronic inflammatory responses. In atopic persons (those prone to allergic reactions), mast cells are "armed" with IgE antibody specific for certain

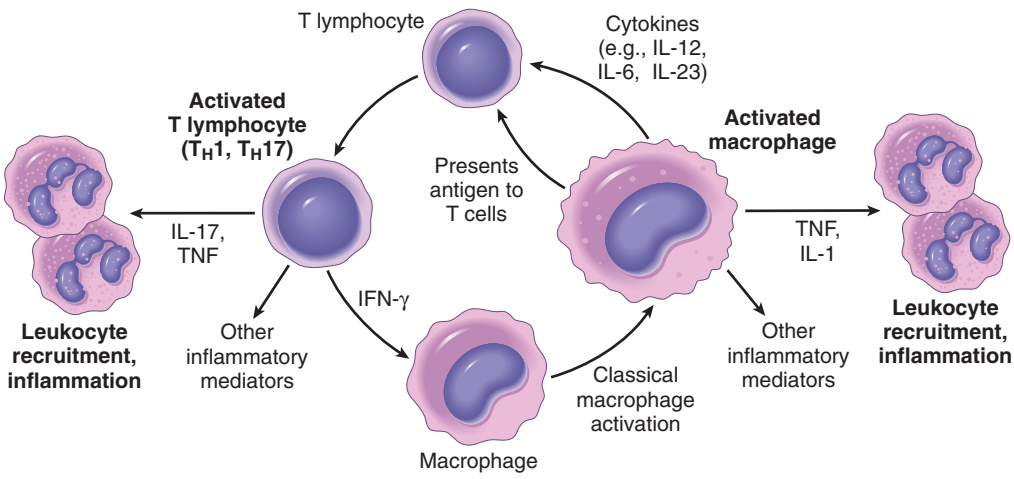


Figure 2–22 Macrophage–lymphocyte interactions in chronic inflammation. Activated lymphocytes and macrophages stimulate each other, and both cell types release inflammatory mediators that affect other cells. IFN-γ, interferon-γ; IL-1, interleukin-1; TNF, tumor necrosis factor.

environmental antigens. When these antigens are subsequently encountered, the IgE-coated mast cells are triggered to release histamines and AA metabolites that elicit the early vascular changes of acute inflammation. IgE-armed mast cells are central players in allergic reactions, including anaphylactic shock (Chapter 4). Mast cells can also elaborate cytokines such as TNF and chemokines and may play a beneficial role in combating some infections.

An important final point: *Although the presence of neutrophils is the hallmark of acute inflammation, many forms of chronic inflammation may continue to show extensive neutrophilic infiltrates,* as a result of either persistent microbes or necrotic cells, or mediators elaborated by macrophages. Such inflammatory lesions are sometimes called "acute on chronic"—for example, in inflammation of bones (osteomyelitis).

Granulomatous Inflammation

Granulomatous inflammation is a distinctive pattern of chronic inflammation characterized by aggregates of activated macrophages with scattered lymphocytes. Granulomas are characteristic of certain specific pathologic states; consequently, recognition of the granulomatous pattern is important because of the limited number of conditions (some life-threatening) that cause it (Table 2–8). Granulomas can form under three settings:

- With persistent T-cell responses to certain microbes (such as *Mycobacterium tuberculosis, T. pallidum,* or fungi), in which T cell–derived cytokines are responsible for chronic macrophage activation. *Tuberculosis is the prototype of a granulomatous disease caused by infection and should always be excluded as the cause when granulomas are identified.*
- Granulomas may also develop in some immune-mediated inflammatory diseases, notably Crohn disease, which is one type of inflammatory bowel disease and an important cause of granulomatous inflammation in the United States.
- They are also seen in a disease of unknown etiology called sarcoidosis, and they develop in response to relatively inert foreign bodies (e.g., suture or splinter), forming so-called *foreign body granulomas.*

The formation of a granuloma effectively "walls off" the offending agent and is therefore a useful defense mechanism. However, granuloma formation does not always lead to eradication of the causal agent, which is frequently resistant to killing or degradation, and granulomatous inflammation with subsequent fibrosis may even be the major cause of organ dysfunction in some diseases, such as tuberculosis.

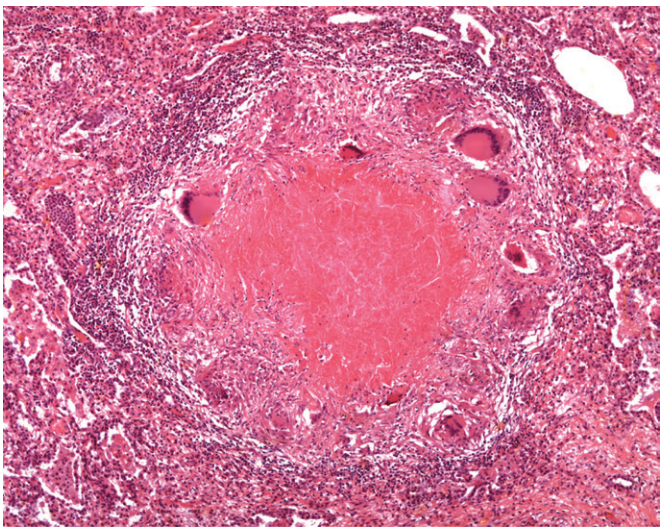


Figure 2–23 A typical granuloma resulting from infection with *Mycobacterium tuberculosis* showing central area of caseous necrosis, activated epithelioid macrophages, giant cells, and a peripheral accumulation of lymphocytes.

MORPHOLOGY

In the usual H&E preparations (Fig. 2–23), some of the activated macrophages in granulomas have pink, granular cytoplasm with indistinct cell boundaries; these are called **epithelioid cells** because of their resemblance to epithelia. Typically, the aggregates of epithelioid macrophages are surrounded by a collar of lymphocytes. Older granulomas may have a rim of fibroblasts and connective tissue. Frequently, but not invariably, multinucleate **giant cells** 40 to 50 μm in diameter are found in granulomas. Such cells consist of a large mass of cytoplasm and many nuclei, and they derive from the fusion of multiple activated macrophages. In granulomas

Table 2–8 Examples of Diseases with Granulomatous Inflammation

Disease	Cause	Tissue Reaction
Tuberculosis	*Mycobacterium tuberculosis*	Caseating granuloma (tubercle): focus of activated macrophages (epithelioid cells), rimmed by fibroblasts, lymphocytes, histiocytes, occasional Langhans giant cells; central necrosis with amorphous granular debris; acid-fast bacilli
Leprosy	*Mycobacterium leprae*	Acid-fast bacilli in macrophages; noncaseating granulomas
Syphilis	*Treponema pallidum*	Gumma: microscopic to grossly visible lesion, enclosing wall of histiocytes; plasma cell infiltrate; central cells are necrotic without loss of cellular outline
Cat-scratch disease	Gram-negative bacillus	Rounded or stellate granuloma containing central granular debris and neutrophils; giant cells uncommon
Sarcoidosis	Unknown etiology	Noncaseating granulomas with abundant activated macrophages
Crohn disease	Immune reaction against intestinal bacteria, self antigens	Occasional noncaseating granulomas in the wall of the intestine, with dense chronic inflammatory infiltrate

associated with certain infectious organisms (most classically the tubercle bacillus), a combination of hypoxia and free radical injury leads to a central zone of necrosis. On gross examination, this has a granular, cheesy appearance and is therefore called **caseous necrosis** (Chapters 1 and 13). On microscopic examination, this necrotic material appears as eosinophilic amorphous, structureless, granular debris, with complete loss of cellular details. The granulomas associated with Crohn disease, sarcoidosis, and foreign body reactions tend to not have necrotic centers and are said to be "non-caseating." Healing of granulomas is accompanied by fibrosis that may be quite extensive.

SUMMARY

Features of Chronic Inflammation

- Prolonged host response to persistent stimulus
- Caused by microbes that resist elimination, immune responses against self and environmental antigens, and some toxic substances (e.g., silica); underlies many important diseases
- Characterized by persistent inflammation, tissue injury, attempted repair by scarring, and immune response
- Cellular infiltrate consisting of activated macrophages, lymphocytes, and plasma cells, often with prominent fibrosis
- Mediated by cytokines produced by macrophages and lymphocytes (notably T lymphocytes), with a tendency to an amplified and prolonged inflammatory response owing to bidirectional interactions between these cells

SYSTEMIC EFFECTS OF INFLAMMATION

Anyone who has suffered a severe bout of viral illness (such as influenza) has experienced the systemic effects of inflammation, collectively called the *acute-phase reaction,* or the systemic inflammatory response syndrome. *The cytokines TNF, IL-1, and IL-6 are the most important mediators of the acute-phase reaction.* These cytokines are produced by leukocytes (and other cell types) in response to infection or in immune reactions and are released systemically. TNF and IL-1 have similar biologic actions, although these may differ in subtle ways (Fig. 2–17). IL-6 stimulates the hepatic synthesis of a number of plasma proteins, described further on.

The acute-phase response consists of several clinical and pathologic changes.

- *Fever,* characterized by an elevation of body temperature, is one of the most prominent manifestations of the acute-phase response. Fever is produced in response to substances called pyrogens that act by stimulating prostaglandin synthesis in the vascular and perivascular cells of the hypothalamus. Bacterial products, such as lipopolysaccharide (LPS) (called *exogenous pyrogens*), stimulate leukocytes to release cytokines such as IL-1 and TNF (called *endogenous pyrogens*), which increase the levels of cyclooxygenases that convert AA into prostaglandins. In the hypothalamus the prostaglandins, especially PGE_2, stimulate the production of neurotransmitters, which function to reset the temperature set point at a higher level. NSAIDs, including aspirin, reduce fever by inhibiting cyclooxygenase and thus blocking prostaglandin synthesis. Although fever was recognized as a sign of infection hundreds of years ago, it is still not clear what the purpose of this reaction may be. An elevated body temperature has been shown to help amphibians ward off microbial infections, and it is assumed that fever does the same for mammals, although the mechanism is unknown.
- *Elevated plasma levels of acute-phase proteins.* These plasma proteins are mostly synthesized in the liver, and in the setting of acute inflammation, their concentrations may increase several hundred-fold. Three of the best known of these proteins are C-reactive protein (CRP), fibrinogen, and serum amyloid A (SAA) protein. Synthesis of these molecules by hepatocytes is stimulated by cytokines, especially IL-6. Many acute-phase proteins, such as CRP and SAA, bind to microbial cell walls, and they may act as opsonins and fix complement, thus promoting the elimination of the microbes. Fibrinogen binds to erythrocytes and causes them to form stacks (rouleaux) that sediment more rapidly at unit gravity than individual erythrocytes. This is the basis for measuring the erythrocyte sedimentation rate (ESR) as a simple test for the systemic inflammatory response, caused by any number of stimuli, including LPS. Serial measurements of ESR and CRP are used to assess therapeutic responses in patients with inflammatory disorders such as rheumatoid arthritis. Elevated serum levels of CRP are now used as a marker for increased risk of myocardial infarction or stroke in patients with atherosclerotic vascular disease. It is believed that inflammation is involved in the development of atherosclerosis (Chapter 9), and increased CRP is a measure of inflammation.
- *Leukocytosis* is a common feature of inflammatory reactions, especially those induced by bacterial infection (see Table 11–6, Chapter 11). The leukocyte count usually climbs to 15,000 to 20,000 cells/mL, but in some extraordinary cases it may reach 40,000 to 100,000 cells/mL. These extreme elevations are referred to as *leukemoid reactions* because they are similar to those seen in leukemia. The leukocytosis occurs initially because of accelerated release of cells (under the influence of cytokines, including TNF and IL-1) from the bone marrow postmitotic reserve pool. Both mature and immature neutrophils may be seen in the blood; the presence of circulating immature cells is referred to as a "shift to the left." Prolonged infection also stimulates production of colony-stimulating factors (CSFs), which increase the bone marrow output of leukocytes, thus compensating for the consumption of these cells in the inflammatory reaction. Most bacterial infections induce an increase in the blood neutrophil count, called neutrophilia. Viral infections, such as infectious mononucleosis, mumps, and German measles, are associated with increased numbers of lymphocytes (lymphocytosis). Bronchial asthma, hay fever, and parasite infestations all involve an increase in the absolute number of eosinophils, creating an

eosinophilia. Certain infections (typhoid fever and infections caused by some viruses, rickettsiae, and certain protozoa) are paradoxically associated with a decreased number of circulating white cells (leukopenia), likely because of cytokine-induced sequestration of lymphocytes in lymph nodes.

- Other manifestations of the acute-phase response include increased heart rate and blood pressure; decreased sweating, mainly as a result of redirection of blood flow from cutaneous to deep vascular beds, to minimize heat loss through the skin; and rigors (shivering), chills (perception of being cold as the hypothalamus resets the body temperature), anorexia, somnolence, and malaise, probably secondary to the actions of cytokines on brain cells.
- In severe bacterial infections (sepsis), the large amounts of bacterial products in the blood or extravascular tissue stimulate the production of several cytokines, notably TNF, as well as IL-12 and IL-1. TNF can cause disseminated intravascular coagulation (DIC), metabolic disturbances including acidosis, and hypotensive shock. This clinical triad is described as *septic shock*; it is discussed in more detail in Chapter 3.

SUMMARY

Systemic Effects of Inflammation

- Fever: cytokines (TNF, IL-1) stimulate production of prostaglandins in hypothalamus
- Production of acute-phase proteins: C-reactive protein, others; synthesis stimulated by cytokines (IL-6, others) acting on liver cells
- Leukocytosis: cytokines (CSFs) stimulate production of leukocytes from precursors in the bone marrow
- In some severe infections, septic shock: fall in blood pressure, disseminated intravascular coagulation, metabolic abnormalities; induced by high levels of TNF

Even before the inflammatory reaction ends, the body begins the process of healing the damage and restoring normal structure and function. This process is called repair, and it involves the proliferation and differentiation of several cell types and the deposition of connective tissue. Defects in tissue repair have serious consequences. Conversely, excessive connective tissue deposition (fibrosis) is also a cause of significant abnormalities. Therefore, the mechanisms and regulation of the repair process are of great physiologic and pathologic importance.

OVERVIEW OF TISSUE REPAIR

Critical to the survival of an organism is the ability to repair the damage caused by toxic insults and inflammation. The inflammatory response to microbes and injured tissues not only serves to eliminate these dangers but also sets into motion the process of repair. Repair, sometimes called

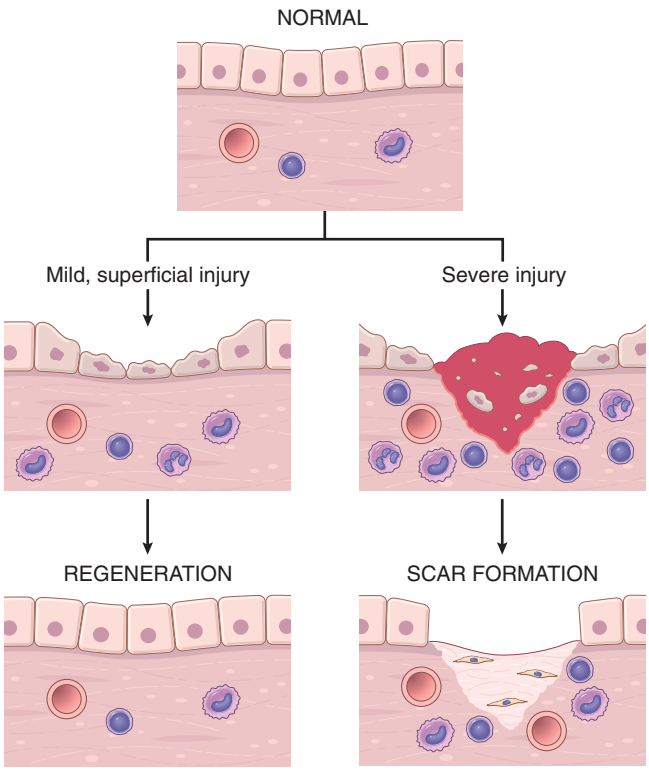


Figure 2–24 Mechanisms of tissue repair: regeneration and scar formation. After mild injury, which damages the epithelium but not the underlying tissue, resolution occurs by regeneration, but after more severe injury with damage to the connective tissue, repair is by scar formation.

healing, refers to the restoration of tissue architecture and function after an injury. It occurs by two types of reactions: regeneration of the injured tissue and scar formation by the deposition of connective tissue (Fig. 2–24).

- *Regeneration.* Some tissues are able to replace the damaged cells and essentially return to a normal state; this process is called regeneration. Regeneration occurs by proliferation of residual (uninjured) cells that retain the capacity to divide, and by replacement from tissue stem cells. It is the typical response to injury in the rapidly dividing epithelia of the skin and intestines, and some parenchymal organs, notably the liver.
- *Scar formation.* If the injured tissues are incapable of regeneration, or if the supporting structures of the tissue are severely damaged, repair occurs by the laying down of connective (fibrous) tissue, a process that results in scar formation. Although the fibrous scar cannot perform the function of lost parenchymal cells, it provides enough structural stability that the injured tissue is usually able to function. The term *fibrosis* is most often used to describe the extensive deposition of collagen that occurs in the lungs, liver, kidney, and other organs as a consequence of chronic inflammation, or in the myocardium after extensive ischemic necrosis (infarction). If fibrosis develops in a tissue space occupied by an inflammatory exudate, it is called organization (as in organizing pneumonia affecting the lung).

After many common types of injury, both regeneration and scar formation contribute in varying degrees to the

ultimate repair. Both processes involve the proliferation of various cells and close interactions between cells and the ECM. The next section discusses the principles of cellular proliferation, the roles of growth factors in the proliferation of different cell types involved in repair, and the roles of stem cells in tissue homeostasis. This is followed by a summary of some important properties of the ECM and how it is involved in repair. These sections lay the foundation for a consideration of the salient features of regeneration and healing by scar formation, concluding with a description of cutaneous wound healing and fibrosis (scarring) in parenchymal organs as illustrations of the repair process.

CELL AND TISSUE REGENERATION

The regeneration of injured cells and tissues involves cell proliferation, which is driven by growth factors and is critically dependent on the integrity of the extracellular matrix. Before describing examples of repair by regeneration, we discuss the general principles of cell proliferation and the functions of the ECM in this process.

The Control of Cell Proliferation

Several cell types proliferate during tissue repair. These include the remnants of the injured tissue (which attempt to restore normal structure), vascular endothelial cells (to create new vessels that provide the nutrients needed for the repair process), and fibroblasts (the source of the fibrous tissue that forms the scar to fill defects that cannot be corrected by regeneration). The proliferation of these cell types is driven by proteins called *growth factors.* The production of polypeptide growth factors and the ability of cells to divide in response to these factors are important determinants of the adequacy of the repair process.

The normal size of cell populations is determined by a balance among cell proliferation, cell death by apoptosis, and emergence of new differentiated cells from stem cells (Fig. 2–25). The key processes in the proliferation of cells are DNA replication and mitosis. The sequence of events that control these two processes is known as the *cell cycle*, described in detail in Chapter 5 in the context of cancer. At this stage, it is sufficient to note that nondividing cells are in cell cycle arrest in the G_1 phase or have exited the cycle and are in the G_0 phase. Growth factors stimulate cells to transition from G_0 into the G_1 phase and beyond into DNA synthesis (S), G_2, and mitosis (M) phases. Progression is regulated by cyclins, whose activity is controlled by cyclin-dependent kinases. Once cells enter the S phase, their DNA is replicated and they progress through G_2 and mitosis.

Proliferative Capacities of Tissues

The ability of tissues to repair themselves is critically influenced by their intrinsic proliferative capacity. On the basis of this criterion, the tissues of the body are divided into three groups.

- *Labile (continuously dividing) tissues.* Cells of these tissues are continuously being lost and replaced by maturation from stem cells and by proliferation of mature cells. Labile cells include hematopoietic cells in the bone marrow and the majority of surface epithelia, such as the stratified squamous surfaces of the skin, oral cavity, vagina, and cervix; the cuboidal epithelia of the ducts draining exocrine organs (e.g., salivary glands, pancreas, biliary tract); the columnar epithelium of the gastrointestinal tract, uterus, and fallopian tubes; and the transitional epithelium of the urinary tract. These tissues can readily regenerate after injury as long as the pool of stem cells is preserved.
- *Stable tissues.* Cells of these tissues are quiescent and have only minimal replicative activity in their normal state. However, these cells are capable of proliferating in response to injury or loss of tissue mass. Stable cells constitute the parenchyma of most solid tissues, such as liver, kidney, and pancreas. They also include endothelial cells, fibroblasts, and smooth muscle cells; the proliferation of these cells is particularly important in wound healing. With the exception of liver, stable tissues have a limited capacity to regenerate after injury.
- *Permanent tissues.* The cells of these tissues are considered to be terminally differentiated and nonproliferative in postnatal life. Most neurons and cardiac muscle cells belong to this category. Thus, injury to brain or heart is irreversible and results in a scar, because neurons and cardiac myocytes cannot regenerate. Limited stem cell replication and differentiation occur in some areas of the adult brain, and there is some evidence that cardiac stem

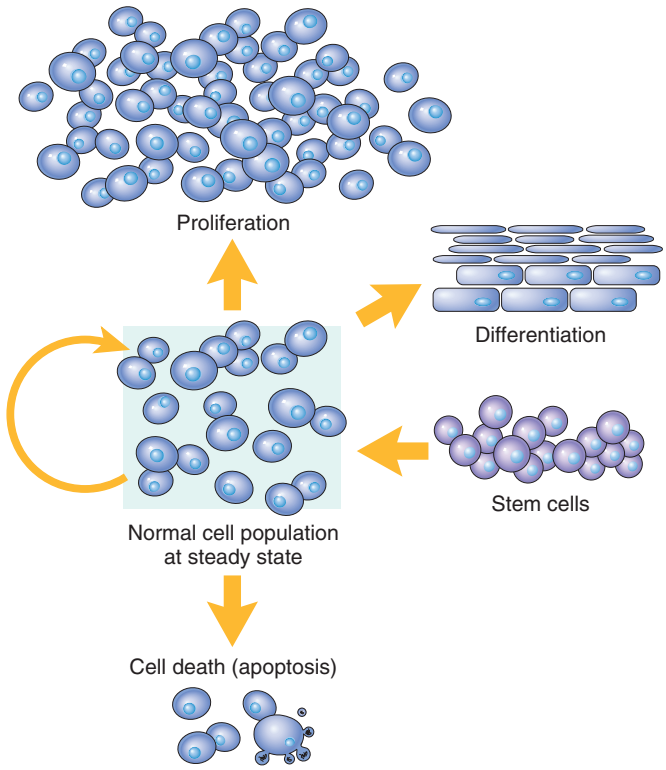


Figure 2–25 Mechanisms regulating cell populations. Cell numbers can be altered by increased or decreased rates of stem cell input, cell death by apoptosis, or changes in the rates of proliferation or differentiation. *(Modified from McCarthy NJ, et al: Apoptosis in the development of the immune system: growth factors, clonal selection and bcl-2. Cancer Metastasis Rev 11:157, 1992.)*

cells may proliferate after myocardial necrosis. Nevertheless, whatever proliferative capacity may exist in these tissues, it is insufficient to produce tissue regeneration after injury. Skeletal muscle is usually classified as a permanent tissue, but satellite cells attached to the endomysial sheath provide some regenerative capacity for this tissue. In permanent tissues, repair is typically dominated by scar formation.

With the exception of tissues composed primarily of nondividing permanent cells (e.g., cardiac muscle, nerve), most mature tissues contain variable proportions of three cell types: continuously dividing cells, quiescent cells that can return to the cell cycle, and cells that have lost replicative ability.

Stem Cells

In most dividing tissues the mature cells are terminally differentiated and short-lived. As mature cells die, the tissue is replenished by the differentiation of cells generated from stem cells. Thus, in these tissues there is a homeostatic equilibrium between the replication, self-renewal, and differentiation of stem cells and the death of the mature, fully differentiated cells. Such relationships are particularly evident in the continuously dividing epithelium of the skin and the gastrointestinal tract, in which stem cells live near the basal layer of the epithelium, and cells differentiate as they migrate to the upper layers of the epithelium before they die and are shed from the surface.

Stem cells are characterized by two important properties: self-renewal capacity and asymmetric replication. Asymmetric replication means that when a stem cell divides, one daughter cell enters a differentiation pathway and gives rise to mature cells, while the other remains an undifferentiated stem cell that retains its self-renewal capacity. Self-renewal enables stem cells to maintain a functional population of precursors for long periods of time. Although the scientific literature is replete with descriptions of various types of stem cells, fundamentally there are two kinds:

- *Embryonic stem cells (ES cells)* are the most undifferentiated stem cells. They are present in the inner cell mass of the blastocyst and have extensive cell renewal capacity. Hence they can be maintained in culture for over a year without differentiating. Under appropriate culture conditions, ES cells can be induced to form specialized cells of all three germ cell layers, including neurons, cardiac muscle, liver cells, and pancreatic islet cells.
- *Adult stem cells,* also called tissue stem cells, are less undifferentiated than ES cells and are found among differentiated cells within an organ or tissue. Although, like ES cells, they also have self-renewal capacity, this property is much more limited. In addition, their lineage potential (ability to give rise to specialized cells) is restricted to some or all of the differentiated cells of the tissue or organ in which they are found.

Whereas the normal function of ES cells is to give rise to all cells of the body, adult stem cells are involved in tissue homeostasis. They maintain the compartment size both in tissues with high turnover, such as skin, bone marrow, and gut epithelium, and in those with low cell turnover, such as heart and blood vessels. Although there is much interest in isolation and infusion of tissue stem cells for replenishment of specialized cells in organs such as the heart (after a myocardial infarct) and brain (after a stroke), tissue stem cells are rare and very difficult to isolate to purity. Furthermore, they occur in specialized microenvironments within the organ called *stem cell niches*. Apparently, signals from other cells in such niches keep the stem cells quiescent and undifferentiated. Stem cell niches have been identified in many organs. In the brain, neural stem cells occur in the subventricular zone and dentate gyrus; in the skin, tissue stem cells are found in the bulge region of the hair follicle; and in the cornea, they are found at the limbus.

Perhaps the most extensively studied tissue stem cells are hematopoietic stem cells found in the bone marrow. Although rare, they can be purified to virtual purity based on cell surface markers. Hematopoietic stem cells can be isolated from bone marrow as well as from the peripheral blood after mobilization by administration of certain cytokines such as granulocyte colony-stimulating factor (G-CSF). As is well known, they can give rise to all blood cell lineages and continuously replenish the formed elements of the blood as these are consumed in the periphery. In clinical practice, marrow stem cells are used for treatment of diseases such as leukemia and lymphomas (Chapter 11). In addition to hematopoietic stem cells, the bone marrow also contains a somewhat distinctive population of tissue stem cells, often called *mesenchymal stem cells*. These cells can give rise to a variety of mesenchymal cells, such as chondroblasts, osteoblasts, and myoblasts. Hence, there is great interest in their therapeutic potential.

The ability to identify and isolate stem cells has given rise to the new field of *regenerative medicine,* which has as its main goal the repopulation of damaged organs by using differentiated progeny of ES cells or adult stem cells. Since ES cells have extensive self-renewal capacity and can give rise to all cell lineages, they often are considered ideal for developing specialized cells for therapeutic purposes. However, since ES cells are derived from blastocysts (typically produced from in vitro fertilization), their progeny carry histocompatibility molecules (human leukocyte antigen [HLA] in people) (Chapter 4) of the donors of the egg and sperm. Thus, they are likely to evoke immunologically mediated rejection by the host, just as organs transplanted from genetically disparate hosts do. Hence, much effort has gone into producing cells with the potential of ES cells from patient tissues. To accomplish this goal, the expressed genes in ES cells and differentiated cells have been compared and a handful of genes that are critical for the "stem-cell-ness" of ES cells have been identified. Introduction of such genes into fully differentiated cells, such as fibroblasts or skin epithelial cells, leads, quite remarkably, to reprogramming of the somatic cell nucleus, such that the cells acquire many of the properties of ES cells. These cells are called *induced pluripotent stem cells (iPS cells)* (Fig. 2–26). Since iPS cells can be derived from each patient, their differentiated progeny should engraft successfully and restore or replace damaged or deficient cells in the patient—for example, insulin-secreting β cells in a patient with diabetes. Although iPS cells hold considerable promise, their clinical usefulness remains to be proved.

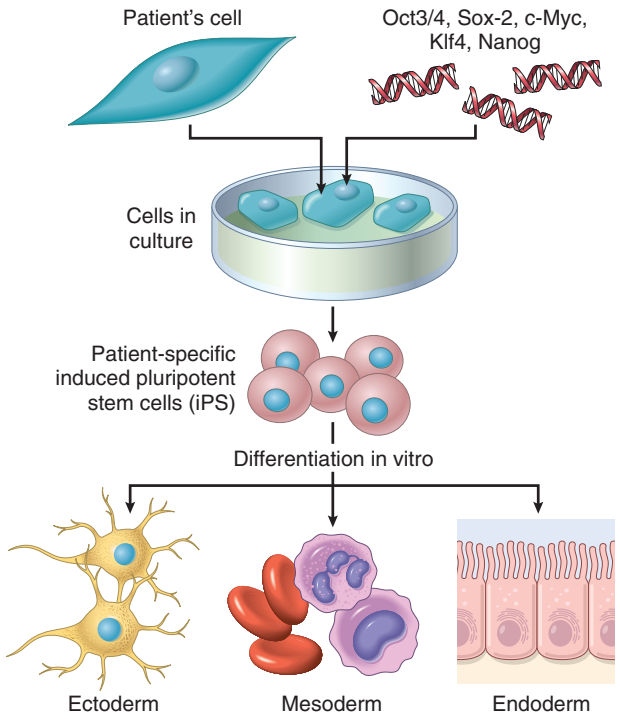


Figure 2–26 The production of induced pluripotent stem cells (iPS cells). Genes that confer stem cell properties are introduced into a patient's differentiated cells, giving rise to stem cells, which can be induced to differentiate into various lineages.

SUMMARY

Cell Proliferation, the Cell Cycle, and Stem Cells

- Regeneration of tissues is driven by proliferation of uninjured (residual) cells and replacement from stem cells.
- Cell proliferation occurs when quiescent cells enter the cell cycle. The cell cycle is tightly regulated by stimulators and inhibitors and contains intrinsic checkpoint controls to prevent replication of abnormal cells.
- Tissues are divided into labile, stable, and permanent, according to the proliferative capacity of their cells.
- Continuously dividing tissues (labile tissues) contain mature cells that are capable of dividing and stem cells that differentiate to replenish lost cells.
- Stem cells from embryos (ES cells) are pluripotent; adult tissues, particularly the bone marrow, contain adult stem cells capable of generating multiple cell lineages.
- Induced pluripotent stem cells (iPS cells) are derived by introducing into mature cells genes that are characteristic of ES cells. iPS cells acquire many characteristics of stem cells.

Growth Factors

Most growth factors are proteins that stimulate the survival and proliferation of particular cells, and may also promote migration, differentiation, and other cellular responses. They induce cell proliferation by binding to specific receptors and affecting the expression of genes whose products typically have several functions: They promote entry of cells into the cell cycle, they relieve blocks on cell cycle progression (thus promoting replication), they prevent apoptosis, and they enhance the synthesis of cellular proteins in preparation for mitosis. A major activity of growth factors is to stimulate the function of growth control genes, many of which are called *proto-oncogenes* because mutations in them lead to unrestrained cell proliferation characteristic of cancer (oncogenesis) (Chapter 5).

There is a huge (and ever-increasing) list of known growth factors. In the following discussion, rather than attempting an exhaustive cataloguing, we highlight only selected molecules that contribute to tissue repair (Table 2–9). Many of the growth factors that are involved in repair are produced by macrophages and lymphocytes that are recruited to the site of injury or are activated at this site, as part of the inflammatory process. Other growth factors are produced by parenchymal cells or stromal (connective tissue) cells in response to cell injury. We start the discussion by describing general principles of growth factor actions. We return to the roles of individual growth factors in the repair process later in the chapter.

Signaling Mechanisms of Growth Factor Receptors

Most growth factors function by binding to specific cell-surface receptors and triggering biochemical signals in cells. The major intracellular signaling pathways induced by growth factor receptors are similar to those of many other cellular receptors that recognize extracellular ligands. In general, these signals lead to the stimulation or repression of gene expression. Signaling may occur directly in the same cell that produces the factor (autocrine signaling), between adjacent cells (paracrine signaling), or over greater distances (endocrine signaling).

Receptor proteins are generally located on the cell surface, but they may be intracellular; in the latter case, the ligands must be sufficiently hydrophobic to enter the cell (e.g., vitamin D, or steroid and thyroid hormones). On the basis of their major signaling transduction pathways, plasma membrane receptors fall into three main types, listed in Table 2–10.

- *Receptors with intrinsic kinase activity.* Binding of ligand to the extracellular portion of the receptor causes dimerization and subsequent phosphorylation of the receptor subunits. Once phosphorylated, the receptors can bind and activate other intracellular proteins (e.g., RAS, phosphatidylinositol 3[PI3]-kinase, phospholipase Cγ [PLC-γ]) and stimulate downstream signals that lead to cell proliferation, or induction of various transcriptional programs.
- *G protein–coupled receptors.* These receptors contain seven-transmembrane α-helix segments and are also known as seven-transmembrane receptors. After ligand binding, the receptors associate with intracellular guanosine triphosphate (GTP)-binding proteins (G proteins) that contain guanosine diphosphate (GDP). Binding of the G proteins causes the exchange of GDP with GTP, resulting in activation of the proteins. Among the several signaling pathways activated through G protein–coupled receptors are those involving cyclic AMP (cAMP), and the generation of inositol 1,4,5-triphosphate (IP_3), which releases calcium from the endoplasmic

Table 2–9 Growth Factors Involved in Regeneration and Repair

Growth Factor	Sources	Functions
Epidermal growth factor (EGF)	Activated macrophages, salivary glands, keratinocytes, and many other cells	Mitogenic for keratinocytes and fibroblasts; stimulates keratinocyte migration; stimulates formation of granulation tissue
Transforming growth factor-α (TGF-α)	Activated macrophages, keratinocytes, many other cell types	Stimulates proliferation of hepatocytes and many other epithelial cells
Hepatocyte growth factor (HGF) (scatter factor)	Fibroblasts, stromal cells in the liver, endothelial cells	Enhances proliferation of hepatocytes and other epithelial cells; increases cell motility
Vascular endothelial growth factor (VEGF)	Mesenchymal cells	Stimulates proliferation of endothelial cells; increases vascular permeability
Platelet-derived growth factor (PDGF)	Platelets, macrophages, endothelial cells, smooth muscle cells, keratinocytes	Chemotactic for neutrophils, macrophages, fibroblasts, and smooth muscle cells; activates and stimulates proliferation of fibroblasts, endothelial, and other cells; stimulates ECM protein synthesis
Fibroblast growth factors (FGFs), including acidic (FGF-1) and basic (FGF-2)	Macrophages, mast cells, endothelial cells, many other cell types	Chemotactic and mitogenic for fibroblasts; stimulates angiogenesis and ECM protein synthesis
Transforming growth factor-β (TGF-β)	Platelets, T lymphocytes, macrophages, endothelial cells, keratinocytes, smooth muscle cells, fibroblasts	Chemotactic for leukocytes and fibroblasts; stimulates ECM protein synthesis; suppresses acute inflammation
Keratinocyte growth factor (KGF) (i.e., FGF-7)	Fibroblasts	Stimulates keratinocyte migration, proliferation, and differentiation

ECM, extracellular membrane.

reticulum. Receptors in this category constitute the largest family of plasma membrane receptors (more than 1500 members have been identified).

- *Receptors without intrinsic enzymatic activity.* These are usually monomeric transmembrane molecules with an extracellular ligand-binding domain; ligand interaction induces an intracellular conformational change that allows association with intracellular protein kinases called Janus kinases (JAKs). Phosphorylation of JAKs activates cytoplasmic transcription factors called STATs (signal transducers and activators of transcription), which shuttle into the nucleus and induce transcription of target genes.

SUMMARY

Growth Factors, Receptors, and Signal Transduction

- Polypeptide growth factors act in autocrine, paracrine, or endocrine manner.
- Growth factors are produced transiently in response to an external stimulus and act by binding to cellular receptors. Different classes of growth factor receptors include receptors with intrinsic kinase activity, G protein–coupled receptors and receptors without intrinsic kinase activity.
- Growth factors such as epidermal growth factor (EGF) and hepatocyte growth factor (HGF) bind to receptors with intrinsic kinase activity, triggering a cascade of phosphorylating events through MAP kinases, which culminate in transcription factor activation and DNA replication.
- G protein–coupled receptors produce multiple effects via the cAMP and Ca^{2+} pathways. Chemokines utilize such receptors.
- Cytokines generally bind to receptors without kinase activity; such receptors interact with cytoplasmic transcription factors that move into the nucleus.
- Most growth factors have multiple effects, such as cell migration, differentiation, stimulation of angiogenesis, and fibrogenesis, in addition to cell proliferation.

Table 2–10 Principal Signaling Pathways Used by Cell Surface Receptors

Receptor Class	Ligands	Signaling Mechanism(s)
Receptors with intrinsic tyrosine kinase activity	EGF, VEGF, FGF, HGF	Ligand binding to one chain of the receptor activates tyrosine kinase on the other chain, resulting in activation of multiple downstream signaling pathways (RAS-MAP kinase, PI-3 kinase, PLC-γ) and activation of various transcription factors.
G protein–coupled seven-transmembrane receptors (GPCRs)	Multiple inflammatory mediators, hormones, all chemokines	Ligand binding induces switch from GDP-bound inactive form of associated G protein to GTP-bound active form; activates cAMP; Ca^{2+} influx leading to increased cell motility; multiple other effects.
Receptors without intrinsic enzymatic activity	Many cytokines including interferons, growth hormone, CSFs, EPO	Ligand binding recruits kinases (e.g., Janus kinases [JAKs]) that phosphorylate and activate transcription factors (e.g., signal transducers and activators of transcription [STATs]).

cAMP, cyclic adenosine monophosphate; CSFs, colony-stimulating factors; EGF, epidermal growth factor; EPO, epopoietin; FGF, fibroblast growth factor; GDP, guanosine diphosphate; GTP, guanosine triphosphate; HGF, hepatocyte growth factor; PI3, phosphatidylinositol-3; PLC-γ, phospholipase Cγ; MAP, microtubule-associated protein; VEGF, vascular endothelial growth factor.

Role of the Extracellular Matrix in Tissue Repair

Tissue repair depends not only on growth factor activity but also on interactions between cells and ECM components. The ECM is a complex of several proteins that assembles into a network that surrounds cells and constitutes a significant proportion of any tissue. *ECM sequesters water, providing turgor to soft tissues, and minerals, giving rigidity to bone. It also regulates the proliferation, movement, and differentiation of the cells living within it, by supplying a substrate for cell adhesion and migration and serving as a reservoir for growth factors.* The ECM is constantly being remodeled; its synthesis and degradation accompany morphogenesis, wound healing, chronic fibrosis, and tumor invasion and metastasis.

ECM occurs in two basic forms: interstitial matrix and basement membrane (Fig. 2–27).

- *Interstitial matrix:* This form of ECM is present in the spaces between cells in connective tissue, and between epithelium and supportive vascular and smooth muscle structures. It is synthesized by mesenchymal cells (e.g., fibroblasts) and tends to form a three-dimensional, amorphous gel. Its major constituents are fibrillar and nonfibrillar collagens, as well as fibronectin, elastin, proteoglycans, hyaluronate, and other elements (described later).
- *Basement membrane:* The seemingly random array of interstitial matrix in connective tissues becomes highly organized around epithelial cells, endothelial cells, and smooth muscle cells, forming the specialized basement membrane. The basement membrane lies beneath the epithelium and is synthesized by overlying epithelium and underlying mesenchymal cells; it tends to form a platelike "chicken wire" mesh. Its major constituents are amorphous nonfibrillar type IV collagen and laminin (see later).

Components of the Extracellular Matrix

There are three basic components of ECM: (1) fibrous structural proteins such as collagens and elastins, which confer tensile strength and recoil; (2) water-hydrated gels such as proteoglycans and hyaluronan, which permit resilience and lubrication; and (3) adhesive glycoproteins that connect the matrix elements to one another and to cells (Fig. 2–27).

Collagen

The collagens are composed of three separate polypeptide chains braided into a ropelike triple helix. Approximately 30 collagen types have been identified, some of which are unique to specific cells and tissues. Some collagen types (e.g., types I, II, III, and V) form fibrils by virtue of lateral cross-linking of the triple helices. The fibrillar collagens form a major proportion of the connective tissue in healing wounds and particularly in scars. The tensile strength of the fibrillar collagens derives from their cross-linking, which is the result of covalent bonds catalyzed by the enzyme lysyl-oxidase. This process is dependent on vitamin C; therefore, individuals with vitamin C deficiency have skeletal deformities, bleed easily because of weak vascular wall basement membrane, and suffer from poor wound healing. Genetic defects in these collagens cause diseases such as osteogenesis imperfecta and Ehlers-Danlos syndrome. Other collagens are nonfibrillar and may form basement membrane (type IV) or be components of other structures such as intervertebral disks (type IX) or dermal–epidermal junctions (type VII).

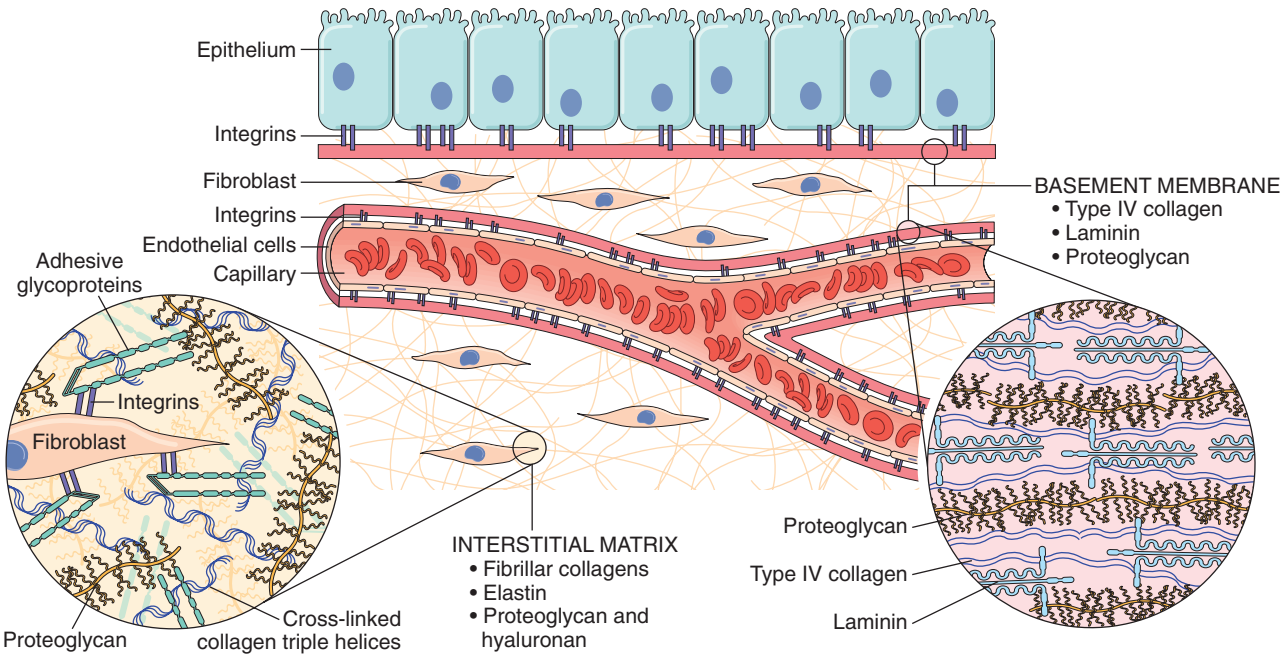


Figure 2–27 The major components of the extracellular matrix (ECM), including collagens, proteoglycans, and adhesive glycoproteins. Note that although there is some overlap in their constituents, basement membrane and interstitial ECM differ in general composition and architecture. Both epithelial and mesenchymal cells (e.g., fibroblasts) interact with ECM through integrins. For simplification, many ECM components have been left out (e.g., elastin, fibrillin, hyaluronan, syndecan).

Elastin

The ability of tissues to recoil and return to a baseline structure after physical stress is conferred by elastic tissue. This is especially important in the walls of large vessels (which must accommodate recurrent pulsatile flow), as well as in the uterus, skin, and ligaments. Morphologically, elastic fibers consist of a central core of elastin surrounded by a meshlike network of fibrillin glycoprotein. Defects in fibrillin synthesis lead to skeletal abnormalities and weakened aortic walls (as in Marfan syndrome, discussed in Chapter 6).

Proteoglycans and Hyaluronan

Proteoglycans form highly hydrated compressible gels conferring resilience and lubrication (such as in the cartilage in joints). They consist of long polysaccharides, called glycosaminoglycans or mucopolysaccharides (examples are dermatan sulfate and heparan sulfate), linked to a protein backbone. Hyaluronan (also called hyaluronic acid), a huge mucopolysaccharide without a protein core, is also an important constituent of the ECM that binds water, and forms a viscous, gelatin-like matrix. Besides providing compressibility to tissues, proteoglycans also serve as reservoirs for growth factors secreted into the ECM (e.g., fibroblast growth factor [FGF], HGF). Some proteoglycans are integral cell membrane proteins that have roles in cell proliferation, migration, and adhesion—for example, by binding growth factors and chemokines and providing high local concentrations of these mediators.

Adhesive Glycoproteins and Adhesion Receptors

Adhesive glycoproteins and adhesion receptors are structurally diverse molecules involved in cell-to-cell adhesion, the linkage of cells to the ECM, and binding between ECM components. The adhesive glycoproteins include fibronectin (a major component of the interstitial ECM) and laminin (a major constituent of basement membrane); they are described here as prototypical of the overall group. The adhesion receptors, also known as cell adhesion molecules (CAMs), are grouped into four families—immunoglobulins, cadherins, selectins, and integrins—of which only the integrins are discussed here.

- *Fibronectin* is a large (450-kDa) disulfide-linked heterodimer synthesized by a variety of cells, including fibroblasts, monocytes, and endothelium that exists in tissue and plasma forms. Fibronectins have specific domains that bind to a wide spectrum of ECM components (e.g., collagen, fibrin, heparin, proteoglycans) and can also attach to cell integrins via a tripeptide arginine–glycine–aspartic acid (abbreviated *RGD*) motif. Tissue fibronectin forms fibrillar aggregates at wound healing sites; plasma fibronectin binds to fibrin within the blood clot that forms in a wound, providing the substratum for ECM deposition and re-epithelialization.
- *Laminin* is the most abundant glycoprotein in basement membrane. It is an 820-kDa cross-shaped heterotrimer that connects cells to underlying ECM components such as type IV collagen and heparan sulfate. Besides mediating attachment to basement membrane, laminin can also modulate cell proliferation, differentiation, and motility.
- *Integrins* are a family of transmembrane heterodimeric glycoprotein chains that were introduced in the context of leukocyte adhesion to endothelium. They are also the main cellular receptors for ECM components, such as fibronectins and laminins. We have already discussed some of the integrins as leukocyte surface molecules that mediate firm adhesion and transmigration across endothelium at sites of inflammation, and we shall meet them again when we discuss platelet aggregation in Chapter 3. Integrins are present in the plasma membrane of most cells, with the exception of red blood cells. They bind to many ECM components through RGD motifs, initiating signaling cascades that can affect cell locomotion, proliferation, and differentiation. Their intracellular domains link to actin filaments, thereby affecting cell shape and mobility.

Functions of the Extracellular Matrix

The ECM is much more than a space filler around cells. Its various functions include

- *Mechanical support* for cell anchorage and cell migration, and maintenance of cell polarity
- *Control of cell proliferation* by binding and displaying growth factors and by signaling through cellular receptors of the integrin family. The type of ECM proteins can affect the degree of differentiation of the cells in the tissue, again acting largely through cell surface integrins.
- *Scaffolding for tissue renewal.* Because maintenance of normal tissue structure requires a basement membrane or stromal scaffold, the integrity of the basement membrane or the stroma of parenchymal cells is critical for the organized regeneration of tissues. Thus, although labile and stable cells are capable of regeneration, disruption of the ECM results in a failure of the tissues to regenerate and repair by scar formation (Fig. 2–24).
- *Establishment of tissue microenvironments.* Basement membrane acts as a boundary between epithelium and underlying connective tissue and also forms part of the filtration apparatus in the kidney.

SUMMARY

Extracellular Matrix and Tissue Repair

- The ECM consists of the *interstitial matrix* between cells, made up of collagens and several glycoproteins, and *basement membranes* underlying epithelia and surrounding vessels, made up of nonfibrillar collagen and laminin.
- The ECM serves several important functions:
  - It provides mechanical support to tissues; this is the role of collagens and elastin.
  - It acts as a substrate for cell growth and the formation of tissue microenvironments.
  - It regulates cell proliferation and differentiation; proteoglycans bind growth factors and display them at high concentration, and fibronectin and laminin stimulate cells through cellular integrin receptors.
- An intact ECM is required for tissue regeneration, and if the ECM is damaged, repair can be accomplished only by scar formation.

Having described the basic components of tissue repair, we now proceed to a discussion of repair by regeneration and by scar formation.

Role of Regeneration in Tissue Repair

The importance of regeneration in the replacement of injured tissues varies in different types of tissues and with the severity of injury.

- In labile tissues, such as the epithelia of the intestinal tract and skin, injured cells are rapidly replaced by proliferation of residual cells and differentiation of tissue stem cells provided the underlying basement membrane is intact. The growth factors involved in these processes are not defined. Loss of blood cells is corrected by proliferation of hematopoietic progenitors in the bone marrow and other tissues, driven by CSFs, which are produced in response to the reduced numbers of blood cells.
- Tissue regeneration can occur in parenchymal organs with stable cell populations, but with the exception of the liver, this is usually a limited process. Pancreas, adrenal, thyroid, and lung have some regenerative capacity. The surgical removal of a kidney elicits in the contralateral kidney a compensatory response that consists of both hypertrophy and hyperplasia of proximal duct cells. The mechanisms underlying this response are not understood.
- The regenerative response of the liver that occurs after surgical removal of hepatic tissue is remarkable and unique among all organs. As much as 40% to 60% of the liver may be removed in a procedure called living-donor transplantation, in which a portion of the liver is resected from a normal person and transplanted into a recipient with end-stage liver disease (Fig. 2–28), or after partial hepatectomy performed for tumor removal. In both situations, the removal of tissue triggers a proliferative response of the remaining hepatocytes (which are normally quiescent), and the subsequent replication of hepatic nonparenchymal cells. In experimental systems, hepatocyte replication after partial hepatectomy is initiated by cytokines (e.g., TNF, IL-6) that prepare the cells for replication by stimulating the transition from G_0 to G_1 in the cell cycle. Progression through the cell cycle is dependent on the activity of growth factors such as HGF (produced by fibroblasts, endothelial cells, and liver nonparenchymal cells) and the EGF family of factors, which includes transforming growth factor-α (TGF-α) (produced by many cell types).

A point worthy of emphasis is that *extensive regeneration or compensatory hyperplasia can occur only if the residual connective tissue framework is structurally intact, as after partial surgical resection. By contrast, if the entire tissue is damaged by infection or inflammation, regeneration is incomplete and is accompanied by scarring*. For example, extensive destruction of the liver with collapse of the reticulin framework, as occurs in a liver abscess, leads to scar formation even though the remaining liver cells have the capacity to regenerate.

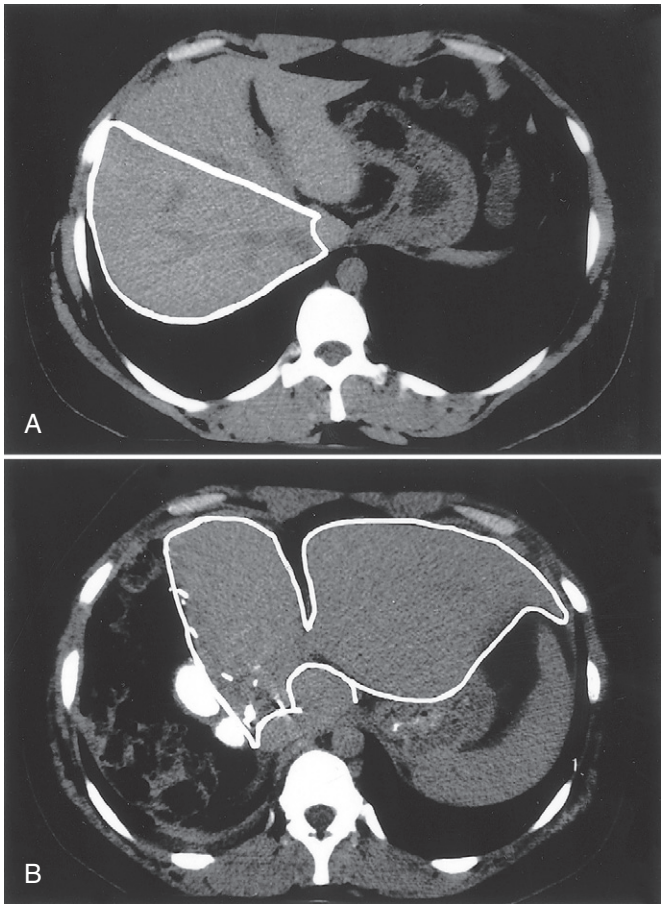


Figure 2–28 Regeneration of the liver. Computed tomography scans show a donor liver in living-donor liver transplantation. **A,** The donor liver before the operation. Note the right lobe (*outline*), which will be resected and used as a transplant. **B,** Scan of the same liver 1 week after resection of the right lobe; note the enlargement of the left lobe (*outline*) without regrowth of the right lobe.
(Courtesy of R. Troisi, MD, Ghent University, Flanders, Belgium.)

SCAR FORMATION

As discussed earlier, if tissue injury is severe or chronic and results in damage to parenchymal cells and epithelia as well as the connective tissue, or if nondividing cells are injured, repair cannot be accomplished by regeneration alone. Under these conditions, repair occurs by replacement of the nonregenerated cells with connective tissue, leading to the formation of a scar, or by a combination of regeneration of some cells and scar formation.

Steps in Scar Formation

Repair by connective tissue deposition consists of sequential processes that follow the inflammatory response (Fig. 2–29):

- Formation of new blood vessels (angiogenesis)
- Migration and proliferation of fibroblasts and deposition of connective tissue, which, together with abundant

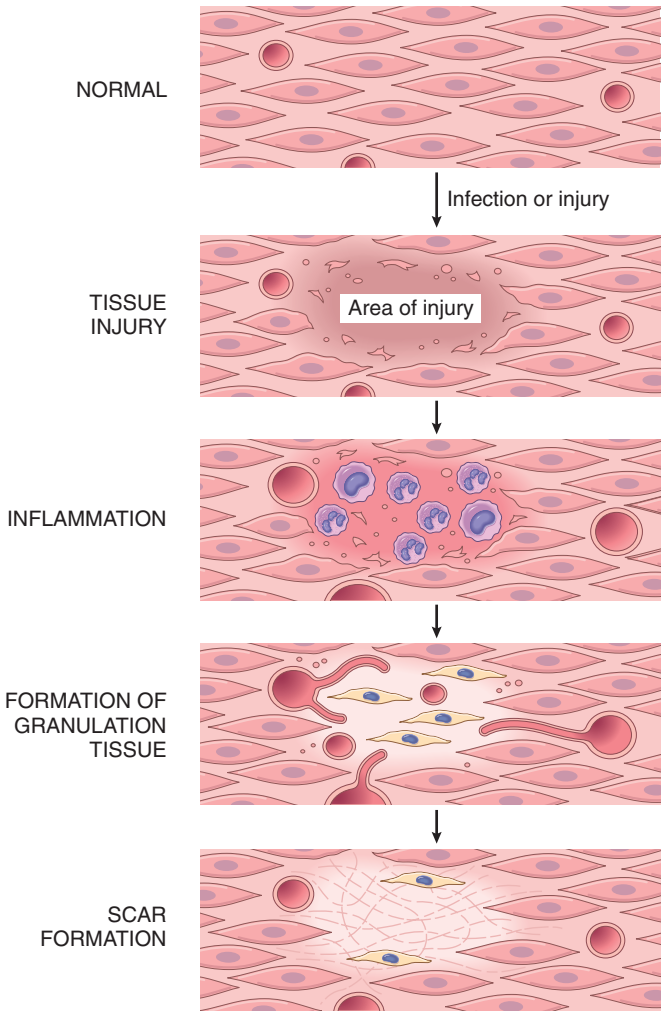


Figure 2–29 Steps in repair by scar formation. Injury to a tissue that has limited regenerative capacity first induces inflammation, which clears dead cells and microbes, if any. This is followed by formation of vascularized granulation tissue and then deposition of ECM to form the scar. ECM, extracellular matrix.

vessels and interspersed leukocytes, has a pink, granular appearance and hence is called *granulation tissue*
- Maturation and reorganization of the fibrous tissue (remodeling) to produce the stable fibrous scar

Repair begins within 24 hours of injury by the emigration of fibroblasts and the induction of fibroblast and endothelial cell proliferation. By 3 to 5 days, the specialized granulation tissue that is characteristic of healing is apparent. The term granulation tissue derives from the gross appearance, such as that beneath the scab of a skin wound. Its histologic appearance is characterized by proliferation of fibroblasts and new thin-walled, delicate capillaries (angiogenesis) in a loose ECM, often with admixed inflammatory cells, mainly macrophages (Fig. 2–30, *A*). Granulation tissue progressively accumulates more fibroblasts, which lay down collagen, eventually resulting in the formation of a scar (Fig. 2–30, *B*). Scars remodel over time. We next describe each of the steps in this process.

Angiogenesis

Angiogenesis is the process of new blood vessel development from existing vessels, primarily venules. It is critical in healing at sites of injury, in the development of collateral circulations at sites of ischemia, and in allowing tumors to increase in size beyond the constraints of their original blood supply. Much work has been done to understand the mechanisms underlying angiogenesis, and therapies to either augment the process (e.g., to improve blood flow to a heart ravaged by coronary atherosclerosis) or inhibit it (e.g., to frustrate tumor growth or block pathologic vessel growth such as in diabetic retinopathy) are being developed.

Angiogenesis involves sprouting of new vessels from existing ones and consists of the following steps (Fig. 2–31):

- Vasodilation occurring in response to NO and increased permeability induced by VEGF
- Separation of pericytes from the abluminal surface
- Migration of endothelial cells toward the area of tissue injury

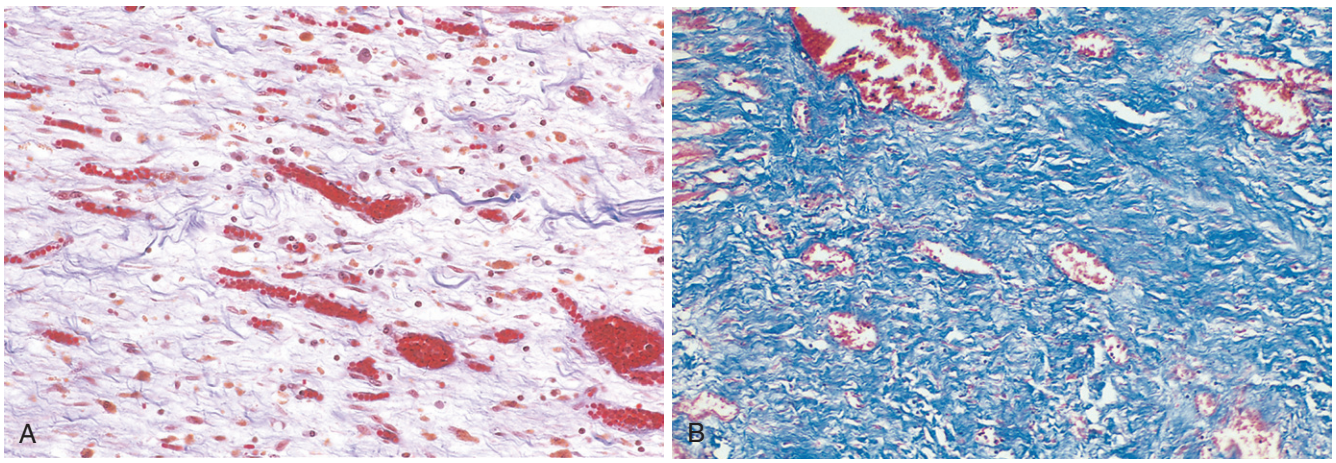


Figure 2–30 A, Granulation tissue showing numerous blood vessels, edema, and a loose ECM containing occasional inflammatory cells. Collagen is stained blue by the trichrome stain; minimal mature collagen can be seen at this point. **B,** Trichrome stain of mature scar, showing dense collagen with only scattered vascular channels. ECM, extracellular matrix.

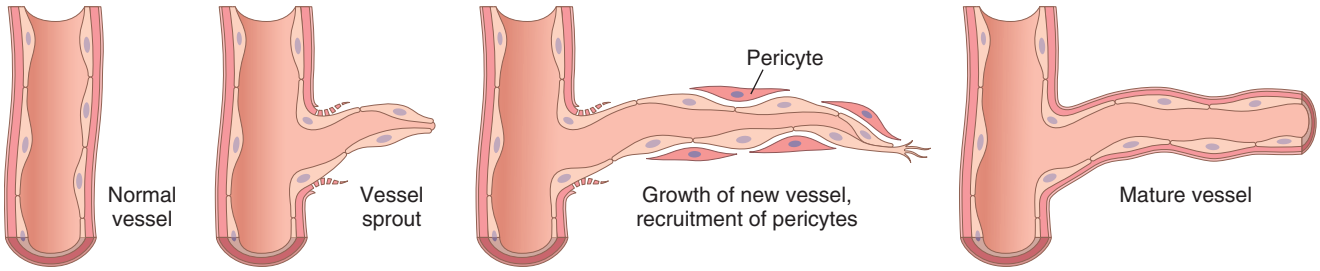


Figure 2–31 Mechanism of angiogenesis. In tissue repair, angiogenesis occurs mainly by growth factor–driven outgrowth of residual endothelium, sprouting of new vessels, and recruitment of pericytes to form new vessels.

- Proliferation of endothelial cells just behind the leading front of migrating cells
- Remodeling into capillary tubes
- Recruitment of periendothelial cells (pericytes for small capillaries and smooth muscle cells for larger vessels) to form the mature vessel
- Suppression of endothelial proliferation and migration and deposition of the basement membrane

The process of angiogenesis involves a variety of growth factors, cell–cell interactions, interactions with ECM proteins, and tissue enzymes.

Growth Factors Involved in Angiogenesis

Several growth factors contribute to angiogenesis; the most important are VEGF and basic fibroblast growth factor (FGF-2).

- The *VEGF family of growth factors* includes VEGF-A, -B, -C, -D, and -E and placental growth factor (PlGF). VEGF-A is generally referred to as VEGF and is the major inducer of angiogenesis after injury and in tumors; VEGF-B and PlGF are involved in vessel development in the embryo; and VEGF-C and -D stimulate both lymphangiogenesis and angiogenesis. VEGFs are expressed in most adult tissues, with the highest expression in epithelial cells adjacent to fenestrated epithelium (e.g., podocytes in the kidney, pigment epithelium in the retina). They bind to a family of tyrosine kinase receptors (VEGFR-1, -2, and -3). The most important of these receptors for angiogenesis is VEGFR-2, which is expressed by VEGF target cells, especially endothelial cells. Of the many inducers of VEGF, hypoxia is the most important; others are platelet-derived growth factor (PDGF), TGF-α, and TGF-β.

  VEGF stimulates both migration and proliferation of endothelial cells, thus initiating the process of capillary sprouting in angiogenesis. It promotes vasodilation by stimulating the production of NO, and contributes to the formation of the vascular lumen. Antibodies against VEGF are approved for the treatment of some tumors that depend on angiogenesis for their spread and growth. These antibodies are also used in the treatment of "wet" (neovascular) age-related macular degeneration, a major cause of visual impairment in adults older than 50 years of age, and is in clinical trials for the treatment of the angiogenesis associated with retinopathy of prematurity and the leaky vessels that lead to diabetic macular edema.
- The *FGF family of growth factors* has more than 20 members; the best characterized are FGF-1 (acidic FGF) and FGF-2 (basic FGF). These growth factors are produced by many cell types and bind to a family of plasma membrane receptors that have tyrosine kinase activity. Released FGF can bind to heparan sulfate and be stored in the ECM. FGF-2 participates in angiogenesis mostly by stimulating the proliferation of endothelial cells. It also promotes the migration of macrophages and fibroblasts to the damaged area, and stimulates epithelial cell migration to cover epidermal wounds.
- *Angiopoietins Ang1 and Ang2* are growth factors that play a role in angiogenesis and the structural maturation of new vessels. Newly formed vessels need to be stabilized by the recruitment of pericytes and smooth muscle cells and by the deposition of connective tissue. Ang1 interacts with a tyrosine kinase receptor on endothelial cells called Tie2. The growth factors PDGF and TGF-β also participate in the stabilization process—PDGF recruits smooth muscle cells and TGF-β suppresses endothelial proliferation and migration, and enhances the production of ECM proteins.

The growth of blood vessels during embryonic development is called *vasculogenesis.* In vasculogenesis, vessels are formed de novo by the coalescence of endothelial precursors called angioblasts. Angioblasts are derived from hemangioblasts, which also provide the precursors of the hematopoietic system. In addition, there are endothelial progenitors in the adult that are derived from bone marrow stem cells and circulate. The contribution of these cells to angiogenesis in adults is not definitely established.

ECM proteins participate in the process of vessel sprouting in angiogenesis, largely through interactions with integrin receptors in endothelial cells and by providing the scaffold for vessel growth. Enzymes in the ECM, notably the matrix metalloproteinases (MMPs), degrade the ECM to permit remodeling and extension of the vascular tube. Newly formed vessels are leaky because of incomplete interendothelial junctions and because VEGF increases vascular permeability. This leakiness explains why granulation tissue is often edematous and accounts in part for the edema that may persist in healing wounds long after the acute inflammatory response has resolved. Furthermore, it leads to high intratumoral pressure and is the basis for the edema that is so problematic in ocular angiogenesis in pathologic processes such as wet macular degeneration.

Activation of Fibroblasts and Deposition of Connective Tissue

The laying down of connective tissue in the scar occurs in two steps: (1) migration and proliferation of fibroblasts into the site of injury and (2) deposition of ECM proteins produced by these cells. The recruitment and activation of fibroblasts to synthesize connective tissue proteins are driven by many growth factors, including PDGF, FGF-2 (described earlier), and TGF-β. The major source of these factors is inflammatory cells, particularly macrophages, which are present at sites of injury and in granulation tissue. Sites of inflammation are also rich in mast cells, and in the appropriate chemotactic milieu, lymphocytes may be present as well. Each of these cell types can secrete cytokines and growth factors that contribute to fibroblast proliferation and activation.

As healing progresses, the number of proliferating fibroblasts and new vessels decreases; however, the fibroblasts progressively assume a more synthetic phenotype, so there is increased deposition of ECM. Collagen synthesis, in particular, is critical to the development of strength in a healing wound site. As described later, collagen synthesis by fibroblasts begins early in wound healing (days 3 to 5) and continues for several weeks, depending on the size of the wound. Net collagen accumulation, however, depends not only on increased synthesis but also on diminished collagen degradation (discussed later). Ultimately, the granulation tissue evolves into a scar composed of largely inactive, spindle-shaped fibroblasts, dense collagen, fragments of elastic tissue, and other ECM components (Fig. 2–30, *B*). As the scar matures, there is progressive vascular regression, which eventually transforms the highly vascularized granulation tissue into a pale, largely avascular scar.

Growth Factors Involved in ECM Deposition and Scar Formation

Many growth factors are involved in these processes, including TGF-β, PDGF, and FGF. Because FGF also is involved in angiogenesis, it was described earlier. Here we briefly describe the major properties of TGF-β and PDGF.

- *Transforming growth factor-β (TGF-β)* belongs to a family of homologous polypeptides (TGF-β1, -β2, and -β3) that includes other cytokines such as bone morphogenetic proteins. The TGF-β1 isoform is widely distributed and is usually referred to as TGF-β. The active factor binds to two cell surface receptors with serine-threonine kinase activity, triggering the phosphorylation of transcription factors called Smads. TGF-β has many and often opposite effects, depending on the cell type and the metabolic state of the tissue. In the context of inflammation and repair, TGF-β has two main functions:
  - TGF-β stimulates the production of collagen, fibronectin, and proteoglycans, and it inhibits collagen degradation by both decreasing proteinase activity and increasing the activity of tissue inhibitors of proteinases known as TIMPs (discussed later on). TGF-β is involved not only in scar formation after injury but also in the development of fibrosis in lung, liver, and kidneys that follows chronic inflammation.
  - TGF-β is an anti-inflammatory cytokine that serves to limit and terminate inflammatory responses. It does so by inhibiting lymphocyte proliferation and the activity of other leukocytes. Mice lacking TGF-β exhibit widespread inflammation and abundant lymphocyte proliferation.
- *Platelet-derived growth factor (PDGF)* belongs to a family of closely related proteins, each consisting of two chains, designated A and B. There are five main PDGF isoforms, of which the BB isoform is the prototype; it is often referred to simply as PDGF. PDGFs bind to receptors designated as PDGFRα and PDGFRβ. PDGF is stored in platelets and released on platelet activation and is also produced by endothelial cells, activated macrophages, smooth muscle cells, and many tumor cells. PDGF causes migration and proliferation of fibroblasts and smooth muscle cells and may contribute to the migration of macrophages.
- *Cytokines* (discussed earlier as mediators of inflammation, and in Chapter 4 in the context of immune responses) may also function as growth factors and participate in ECM deposition and scar formation. IL-1 and IL-13, for example, act on fibroblasts to stimulate collagen synthesis, and can also enhance the proliferation and migration of fibroblasts.

Remodeling of Connective Tissue

After its synthesis and deposition, the connective tissue in the scar continues to be modified and remodeled. Thus, the outcome of the repair process is a balance between synthesis and degradation of ECM proteins. We have already discussed the cells and factors that regulate ECM synthesis. *The degradation of collagens and other ECM components is accomplished by a family of matrix metalloproteinases (MMPs),* which are dependent on zinc ions for their activity. MMPs should be distinguished from neutrophil elastase, cathepsin G, plasmin, and other serine proteinases that can also degrade ECM but are not metalloenzymes. MMPs include interstitial collagenases, which cleave fibrillar collagen (MMP-1, -2, and -3); gelatinases (MMP-2 and -9), which degrade amorphous collagen and fibronectin; and stromelysins (MMP-3, -10, and -11), which degrade a variety of ECM constituents, including proteoglycans, laminin, fibronectin, and amorphous collagen.

MMPs are produced by a variety of cell types (fibroblasts, macrophages, neutrophils, synovial cells, and some epithelial cells), and their synthesis and secretion are regulated by growth factors, cytokines, and other agents. The activity of the MMPs is tightly controlled. They are produced as inactive precursors (zymogens) that must be first activated; this is accomplished by proteases (e.g., plasmin) likely to be present only at sites of injury. In addition, activated MMPs can be rapidly inhibited by specific tissue inhibitors of metalloproteinases (TIMPs), produced by most mesenchymal cells. Thus, during scarring, MMPs are activated to remodel the deposited ECM, and then their activity is shut down by the TIMPs.

SUMMARY

Repair by Scar Formation

- Tissues can be repaired by regeneration with complete restoration of form and function, or by replacement with connective tissue and scar formation.
- Repair by connective tissue deposition involves angiogenesis, migration and proliferation of fibroblasts, collagen synthesis, and connective tissue remodeling.
- Repair by connective tissue starts with the formation of granulation tissue and culminates in the laying down of fibrous tissue.
- Multiple growth factors stimulate the proliferation of the cell types involved in repair.
- TGF-β is a potent fibrogenic agent; ECM deposition depends on the balance among fibrogenic agents, the metalloproteinases (MMPs) that digest ECM, and the TIMPs.

FACTORS THAT INFLUENCE TISSUE REPAIR

Tissue repair may be altered by a variety of influences, frequently reducing the quality or adequacy of the reparative process. Variables that modify healing may be extrinsic (e.g., infection) or intrinsic to the injured tissue. Particularly important are infections and diabetes.

- *Infection* is clinically the most important cause of delay in healing; it prolongs inflammation and potentially increases the local tissue injury.
- *Nutrition* has profound effects on repair; protein deficiency, for example, and especially vitamin C deficiency inhibit collagen synthesis and retard healing.
- *Glucocorticoids* (steroids) have well-documented anti-inflammatory effects, and their administration may result in weakness of the scar because of inhibition of TGF-β production and diminished fibrosis. In some instances, however, the anti-inflammatory effects of glucocorticoids are desirable. For example, in corneal infections, glucocorticoids are sometimes prescribed (along with antibiotics) to reduce the likelihood of opacity that may result from collagen deposition.
- *Mechanical variables* such as increased local pressure or torsion may cause wounds to pull apart, or dehisce.
- *Poor perfusion,* due either to arteriosclerosis and diabetes or to obstructed venous drainage (e.g., in varicose veins), also impairs healing.
- *Foreign bodies* such as fragments of steel, glass, or even bone impede healing.
- The type and extent of tissue injury affects the subsequent repair. Complete restoration can occur only in tissues composed of stable and labile cells; injury to tissues composed of permanent cells must inevitably result in scarring, as in healing of a myocardial infarct.
- The *location of the injury* and the character of the tissue in which the injury occurs are also important. For example, inflammation arising in tissue spaces (e.g., pleural, peritoneal, or synovial cavities) develops extensive exudates. Subsequent repair may occur by digestion of the exudate, initiated by the proteolytic enzymes of leukocytes and resorption of the liquefied exudate. This is called resolution, and generally, in the absence of cellular necrosis, normal tissue architecture is restored. In the setting of larger accumulations, however, the exudate undergoes organization: Granulation tissue grows into the exudate, and a fibrous scar ultimately forms.
- *Aberrations of cell growth* and ECM production may occur even in what begins as normal wound healing. For example, the accumulation of exuberant amounts of collagen can give rise to prominent, raised scars known as *keloids* (Fig. 2–32). There appears to be a heritable predisposition to keloid formation, and the condition is more common in African-Americans. Healing wounds may also generate exuberant granulation tissue that protrudes above the level of the surrounding skin and hinders re-epithelialization. Such tissue is called "proud

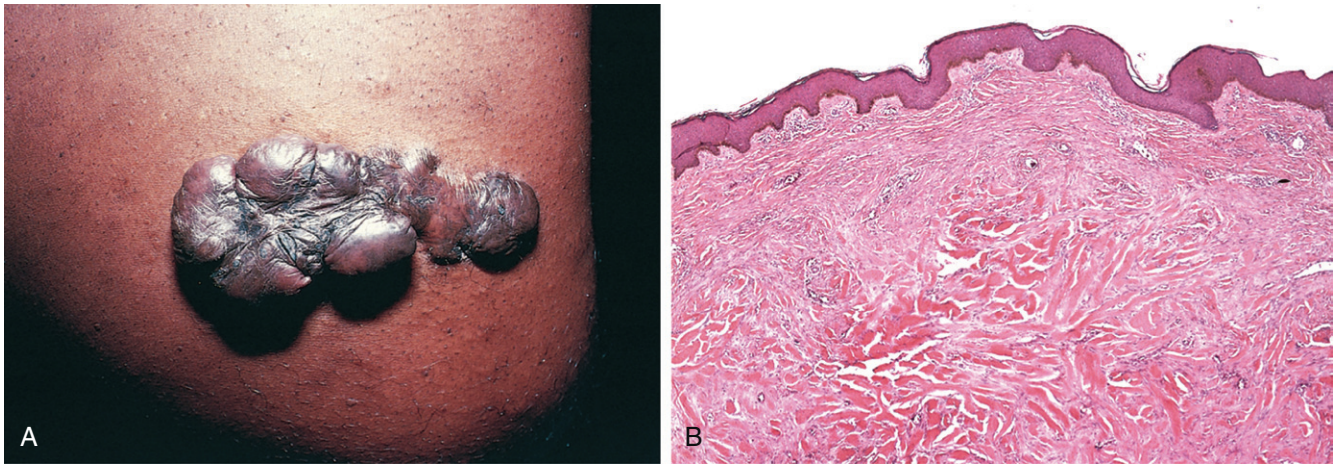


Figure 2–32 Keloid. **A,** Excess collagen deposition in the skin forming a raised scar known as a keloid. **B,** Thick connective tissue deposition in the dermis.

(A, From Murphy GF, Herzberg AJ: Atlas of Dermatology. Philadelphia, WB Saunders, 1996. B, Courtesy of Z. Argenyi, MD, University of Washington, Seattle, Washington.)

flesh" in old medical parlance, and restoration of epithelial continuity requires cautery or surgical resection of the granulation tissue.

SELECTED CLINICAL EXAMPLES OF TISSUE REPAIR AND FIBROSIS

Thus far we have discussed the general principles and mechanisms of repair by regeneration and scarring. In this section we describe two clinically significant types of repair—the healing of skin wounds (cutaneous wound healing) and fibrosis in injured parenchymal organs.

Healing of Skin Wounds

Cutaneous wound healing is a process that involves both epithelial regeneration and the formation of connective tissue scar and is thus illustrative of the general principles that apply to healing in all tissues.

Depending on the nature and size of the wound, the healing of skin wounds is said to occur by first or second intention.

Healing by First Intention

One of the simplest examples of wound repair is the healing of a clean, uninfected surgical incision approximated by surgical sutures (Fig. 2–33). This is referred to as primary

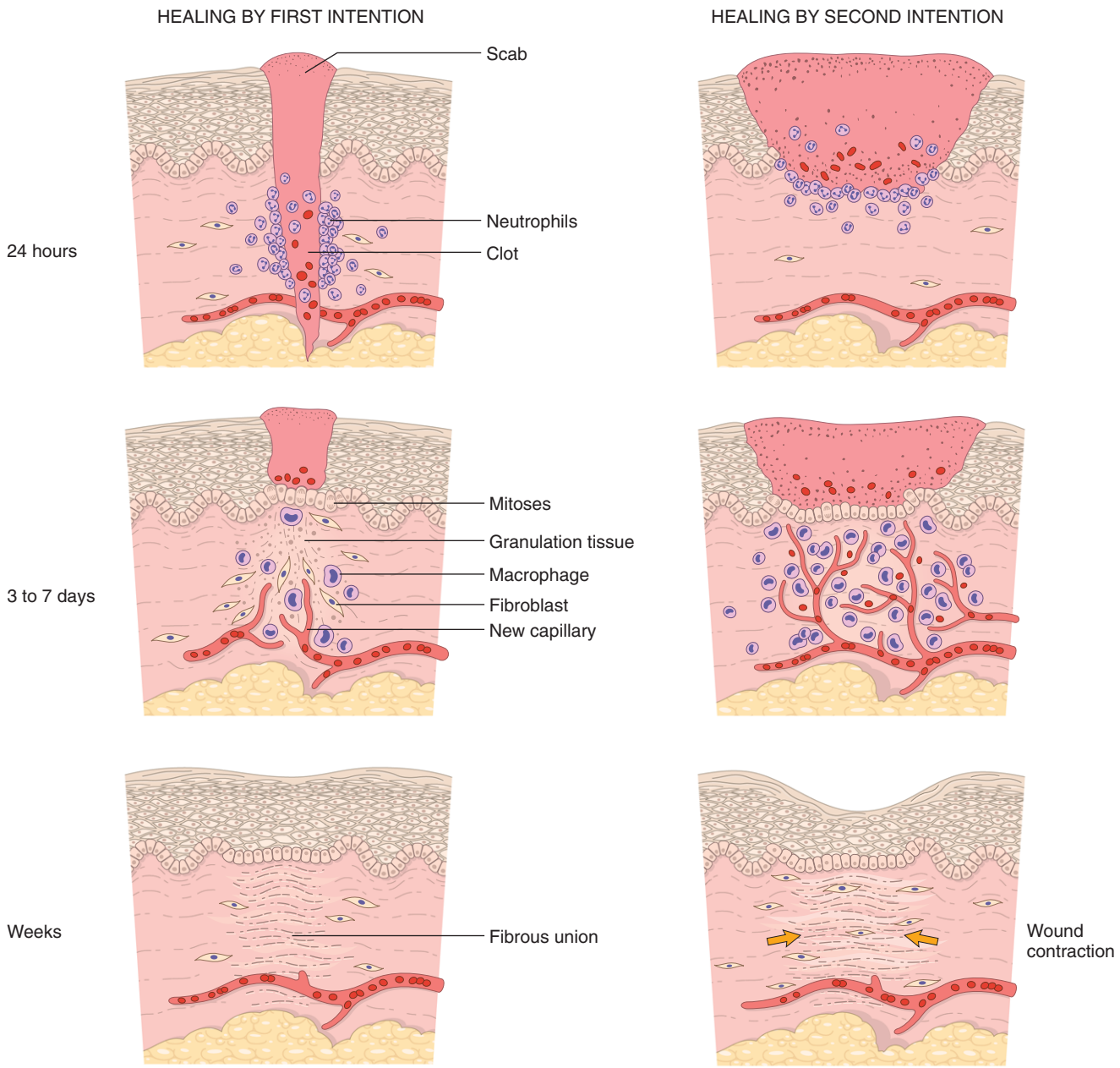


Figure 2–33 Steps in wound healing by first intention (*left*) and second intention (*right*). In the latter case, note the large amount of granulation tissue and wound contraction.

union, or healing by first intention. The incision causes only focal disruption of epithelial basement membrane continuity and death of relatively few epithelial and connective tissue cells. As a result, *epithelial regeneration is the principal mechanism of repair*. A small scar is formed, but there is minimal wound contraction. The narrow incisional space first fills with fibrin-clotted blood, which then is rapidly invaded by granulation tissue and covered by new epithelium. The steps in the process are well defined:

- Within 24 hours, neutrophils are seen at the incision margin, migrating toward the fibrin clot. Basal cells at the cut edge of the epidermis begin to show increased mitotic activity. Within 24 to 48 hours, epithelial cells from both edges have begun to migrate and proliferate along the dermis, depositing basement membrane components as they progress. The cells meet in the midline beneath the surface scab, yielding a thin but continuous epithelial layer.
- By day 3, neutrophils have been largely replaced by macrophages, and granulation tissue progressively invades the incision space. Collagen fibers are now evident at the incision margins, but these are vertically oriented and do not bridge the incision. Epithelial cell proliferation continues, yielding a thickened epidermal covering layer.
- By day 5, neovascularization reaches its peak as granulation tissue fills the incisional space. Collagen fibrils become more abundant and begin to bridge the incision. The epidermis recovers its normal thickness as differentiation of surface cells yields a mature epidermal architecture with surface keratinization.
- During the second week, there is continued collagen accumulation and fibroblast proliferation. The leukocyte infiltrate, edema, and increased vascularity are substantially diminished. The long process of "blanching" begins, accomplished by increasing collagen deposition within the incisional scar and the regression of vascular channels.
- By the end of the first month, the scar consists of a cellular connective tissue, largely devoid of inflammatory cells, covered by an essentially normal epidermis. However, the dermal appendages destroyed in the line of the incision are permanently lost. The tensile strength of the wound increases with time, as described later.

Healing by Second Intention

When cell or tissue loss is more extensive, such as in large wounds, at sites of abscess formation, ulceration, and ischemic necrosis (infarction) in parenchymal organs, the repair process is more complex and involves a combination of regeneration and scarring. In second intention healing of skin wounds, also known as healing by secondary union (Fig. 2–34; see also Fig. 2–33), the inflammatory reaction is

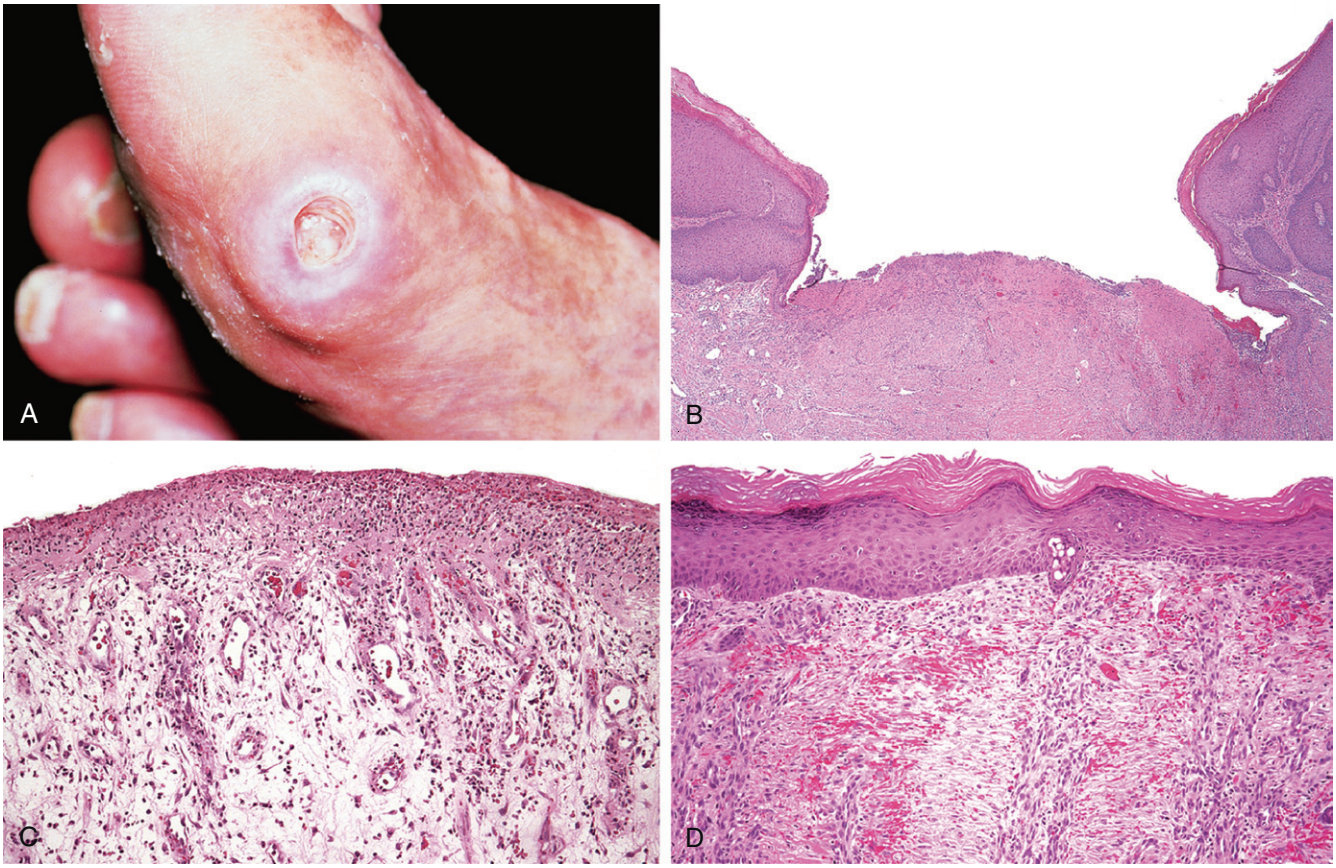


Figure 2–34 Healing of skin ulcers. **A,** Pressure ulcer of the skin, commonly found in diabetic patients. **B,** A skin ulcer with a large gap between the edges of the lesion. **C,** A thin layer of epidermal re-epithelialization, and extensive granulation tissue formation in the dermis. **D,** Continuing re-epithelialization of the epidermis and wound contraction.
(Courtesy of Z. Argenyi, MD, University of Washington, Seattle, Wash.)

more intense, and there is development of abundant granulation tissue, with accumulation of ECM and formation of a large scar, followed by wound contraction mediated by the action of myofibroblasts.

Secondary healing differs from primary healing in several respects:

- A larger clot or scab rich in fibrin and fibronectin forms at the surface of the wound.
- Inflammation is more intense because large tissue defects have a greater volume of necrotic debris, exudate, and fibrin that must be removed. Consequently, large defects have a greater potential for secondary, inflammation-mediated, injury.
- Larger defects require a greater volume of granulation tissue to fill in the gaps and provide the underlying framework for the regrowth of tissue epithelium. A greater volume of granulation tissue generally results in a greater mass of scar tissue.
- Secondary healing involves wound contraction. Within 6 weeks, for example, large skin defects may be reduced to 5% to 10% of their original size, largely by contraction. This process has been ascribed to the presence of myofibroblasts, which are modified fibroblasts exhibiting many of the ultrastructural and functional features of contractile smooth muscle cells.

Wound Strength

Carefully sutured wounds have approximately 70% of the strength of normal skin, largely because of the placement of sutures. When sutures are removed, usually at 1 week, wound strength is approximately 10% of that of unwounded skin, but this increases rapidly over the next 4 weeks. The recovery of tensile strength results from collagen synthesis exceeding degradation during the first 2 months, and from structural modifications of collagen (e.g., cross-linking, increased fiber size) when synthesis declines at later times. Wound strength reaches approximately 70% to 80% of normal by 3 months and usually does not improve substantially beyond that point.

Fibrosis in Parenchymal Organs

Deposition of collagen is part of normal wound healing. The term *fibrosis* is used to denote the excessive deposition of collagen and other ECM components in a tissue. As already mentioned, the terms *scar* and *fibrosis* are used interchangeably, but *fibrosis* most often refers to the deposition of collagen in chronic diseases.

The basic mechanisms of fibrosis are the same as those of scar formation during tissue repair. However, tissue repair typically occurs after a short-lived injurious stimulus and follows an orderly sequence of steps, whereas fibrosis is induced by persistent injurious stimuli such as infections, immunologic reactions, and other types of tissue injury. The fibrosis seen in chronic diseases such as pulmonary fibrosis is often responsible for organ dysfunction and even organ failure.

SUMMARY

Cutaneous Wound Healing and Pathologic Aspects of Repair

- Cutaneous wounds can heal by primary union (first intention) or secondary union (second intention); secondary healing involves more extensive scarring and wound contraction.
- Wound healing can be altered by many conditions, particularly infection and diabetes; the type, volume, and location of the injury are also important factors in healing.
- Excessive production of ECM can cause keloids in the skin.
- Persistent stimulation of collagen synthesis in chronic inflammatory diseases leads to fibrosis of the tissue.

BIBLIOGRAPHY

Bradley JR: TNF-mediated inflammatory disease. J Pathol 214:149, 2008. *[An overview of the biology of TNF and the clinical utility of TNF antagonists.]*

Carlson BM: Some principles of regeneration in mammalian systems. Anat Rec 287:4, 2005. *[A thoughtful review of the evolutionary aspects and general mechanisms of limb and organ regeneration.]*

Carmeliet P: Angiogenesis in life, disease and medicine. Nature 438:932, 2005. *[A review of the main aspects of normal and abnormal angiogenesis.]*

Charo IF, Ransohoff RM: The many roles of chemokines and chemokine receptors in inflammation. N Engl J Med 354:610, 2006. *[An overview of the functions of chemokines in inflammation.]*

Fausto N: Liver regeneration and repair: hepatocytes, progenitor cells and stem cells. Hepatology 39:1477, 2004. *[A review of the cellular and molecular mechanisms of liver regeneration.]*

Gabay C, Lamacchia C, Palmer G: IL-1 pathways in inflammation and human diseases. Nat Rev Rheumatol 6:232, 2010. *[An excellent review of the biology of IL-1 and the therapeutic targeting of this cytokine in inflammatory diseases.]*

Gurtner GC, Werner S, Barrandon Y, Longaker MT: Wound repair and regeneration. Nature 453:314, 2008. *[An excellent review of the principles of tissue regeneration and repair.]*

Hynes RO: Integrins: bidirectional, allosteric signaling machines. Cell 110:673, 2002. *[An excellent review of the molecular mechanisms of integrin signaling, linking ECM components to intracellular signal transduction pathways.]*

Jiang D, Liang J, Noble PW: Hyaluronans in tissue injury and repair. Annu Rev Cell Dev Biol 23:435, 2007. *[A discussion of the role of a major family of ECM proteins in tissue repair.]*

Khanapure SP, Garvey DS, Janero DR, et al: Eicosanoids in inflammation: biosynthesis, pharmacology, and therapeutic frontiers. Curr Top Med Chem 7:311, 2007. *[A summary of the properties of this important class of inflammatory mediators.]*

Lentsch AB, Ward PA: Regulation of inflammatory vascular damage. J Pathol 190:343, 2000. *[Discussion of the mechanisms of endothelial damage and increased vascular permeability.]*

Ley K, Laudanna C, Cybulsky MI, Nourshargh S: Getting to the site of inflammation: the leukocyte adhesion cascade updated. Nat Rev Immunol 7:678, 2007. *[A modern discussion of leukocyte recruitment to sites of inflammation.]*

Martin P, Leibovich SJ: Inflammatory cells during wound repair: the good, the bad, and the ugly. Trends Cell Biol 15:599, 2005. *[Good review on the multiple roles of inflammatory cells in repair.]*

Masters SL, Simon A, Aksentijevich I, Kastner DL: Horror autoinflammaticus: the molecular pathophysiology of autoinflammatory disease. Annu Rev Immunol 27:621, 2009. *[An excellent discussion of autoinflammatory syndromes caused by gain-of-function mutations in components of the inflammasome.]*

McAnulty RJ: Fibroblasts and myofibroblasts: their source, function, and role in disease. Int J Biochem Cell Biol 39:666, 2007. *[A discussion*

of the two major types of stroma cells and their roles on tissue repair and fibrosis.]

Muller WA: Mechanisms of leukocyte transendothelial migration. Annu Rev Pathol 6:323, 2011. *[A thoughtful review of the mechanisms by which leukocytes traverse the endothelium.]*

Nagy JA, Dvorak AM, Dvorak HF: VEGF-A and the induction of pathological angiogenesis. Annu Rev Pathol 2:251, 2007. *[A review of the VEGF family of growth factors and their role in angiogenesis in cancer, inflammation, and various disease states.]*

Nathan C, Ding A: Nonresolving inflammation. Cell 140:871, 2010. *[A discussion of the abnormalities that lead to chronic inflammation.]*

Page-McCaw A, Ewald AJ, Werb Z: Matrix metalloproteinases and the regulation of tissue remodelling. Nat Rev Mol Cell Biol 8:221, 2007. *[A review of the function of matrix modifying enzymes in tissue repair.]*

Papayannapoulos V, Zychlinsky A: NETs: a new strategy for using old weapons. Trends Immunol 30:513, 2009. *[A review of a newly discovered mechanism by which neutrophils destroy microbes.]*

Ricklin D, Hajishengallis G, Yang K, Lambris JD: Complement: a key system for immune surveillance and homeostasis. Nat Immunol 11:785, 2010. *[A current overview of the activation and functions of the complement system and its role in disease.]*

Rock KL, Kono H: The inflammatory response to cell death. Annu Rev Pathol 3:99, 2008. *[An excellent discussion of how the immune system recognizes necrotic cells.]*

Schultz GS, Wysocki A: Interactions between extracellular matrix and growth factors in wound healing. Wound Repair Regen 17:153, 2009. *[A discussion of the regulation of growth factors by the ECM.]*

Schroder K, Tschopp J: The inflammasomes. Cell 140:821, 2010. *[An excellent review of the cellular machinery that recognizes products of dead cells, many foreign and abnormal substances, and some microbes.]*

Segal AW: How neutrophils kill microbes. Annu Rev Immunol 23:197, 2005. *[An excellent discussion of the microbicidal mechanisms of neutrophils.]*

Stappenbeck TS, Miyoshi H: The role of stromal stem cells in tissue regeneration and wound repair. Science 324:1666, 2009. *[An excellent review of the role of tissue stem cells in repair.]*

Stearns-Kurosawa DJ, Osuchowski MF, Valentine C, et al: The pathogenesis of sepsis. Annu Rev Pathol 6:19, 2011. *[A discussion of the current concepts of pathogenic mechanisms in sepsis and septic shock.]*

Takeuchi O, Akira S: Pattern recognition receptors and inflammation. Cell 140:805, 2010. *[An excellent overview of Toll-like receptors and other pattern recognition receptor families, and their roles in host defense and inflammation.]*

Wynn TA: Cellular and molecular mechanisms of fibrosis. J Pathol 214:199, 2008. *[An overview of the cellular mechanisms of fibrosis, with an emphasis on the role of the immune system in fibrotic reactions to chronic infections.]*

Yamanaka S, Blau HM: Nuclear reprogramming to a pluripotent state by three approaches. Nature 465:704, 2010. *[A review of the exciting technology for generating iPS cells for regenerative medicine.]*

This page intentionally left blank

See Targeted Therapy available online at
studentconsult.com

CHAPTER 3

Hemodynamic Disorders, Thromboembolism, and Shock

CHAPTER CONTENTS

The health of cells and tissues depends on the circulation of blood, which delivers oxygen and nutrients and removes wastes generated by cellular metabolism. Under normal conditions, as blood passes through capillary beds, proteins in the plasma are retained within the vasculature and there is little net movement of water and electrolytes into the tissues. This balance is often disturbed by pathologic conditions that alter endothelial function, increase vascular pressure, or decrease plasma protein content, all of which promote *edema*—accumulation of fluid resulting from a net outward movement of water into extravascular spaces. Depending on its severity and location, edema may have minimal or profound effects. In the lower extremities, it may only make one's shoes feel snugger after a long sedentary day; in the lungs, however, edema fluid can fill alveoli, causing life-threatening hypoxia.

Our blood vessels are frequently subject to trauma of varying degrees. *Hemostasis* is the process of blood clotting that prevents excessive bleeding after blood vessel damage. Inadequate hemostasis may result in *hemorrhage,* which can compromise regional tissue perfusion and, if massive and rapid, may lead to *hypotension, shock,* and death. Conversely, inappropriate clotting (*thrombosis*) or migration of clots (*embolism*) can obstruct blood vessels, potentially causing ischemic cell death (*infarction*). Indeed, *thromboembolism* lies at the heart of three major causes of morbidity and death in developed countries: myocardial infarction, pulmonary embolism, and cerebrovascular accident (stroke).

HYPEREMIA AND CONGESTION

Hyperemia and congestion both refer to an increase in blood volume within a tissue but they have different underlying mechanisms. *Hyperemia is an active process* resulting from arteriolar dilation and increased blood inflow, as occurs at sites of inflammation or in exercising skeletal muscle. Hyperemic tissues are redder than normal because of engorgement with oxygenated blood. *Congestion is a passive process* resulting from impaired outflow of venous blood from a tissue. It can occur systemically, as in cardiac failure, or locally as a consequence of an isolated venous obstruction. Congested tissues have an abnormal blue-red color (*cyanosis*) that stems from the accumulation of deoxygenated hemoglobin in the affected area. In long-standing *chronic congestion,* inadequate tissue perfusion and persistent hypoxia may lead to parenchymal cell death and secondary tissue fibrosis, and the elevated intravascular pressures may cause edema or sometimes rupture capillaries, producing focal hemorrhages.

MORPHOLOGY

Cut surfaces of hyperemic or congested tissues feel wet and typically ooze blood. On microscopic examination, **acute pulmonary congestion** is marked by blood-engorged alveolar capillaries and variable degrees of alveolar septal edema and intra-alveolar hemorrhage. In **chronic pulmonary congestion,** the septa become thickened and fibrotic, and the alveolar spaces contain numerous macrophages laden with hemosiderin ("heart failure cells") derived from phagocytosed red cells. In **acute hepatic congestion,** the central vein and sinusoids are distended with blood, and there may even be central hepatocyte dropout due to necrosis. The periportal hepatocytes, better oxygenated because of their proximity to hepatic arterioles, experience less severe hypoxia and may develop only reversible fatty change. In **chronic passive congestion of the liver,** the central regions of the hepatic lobules, viewed on gross examination, are red-brown and slightly depressed (owing to cell loss) and are accentuated against the surrounding zones of uncongested tan, sometimes fatty, liver **(nutmeg liver)** (Fig. 3–1, *A*). Microscopic findings include centrilobular hepatocyte

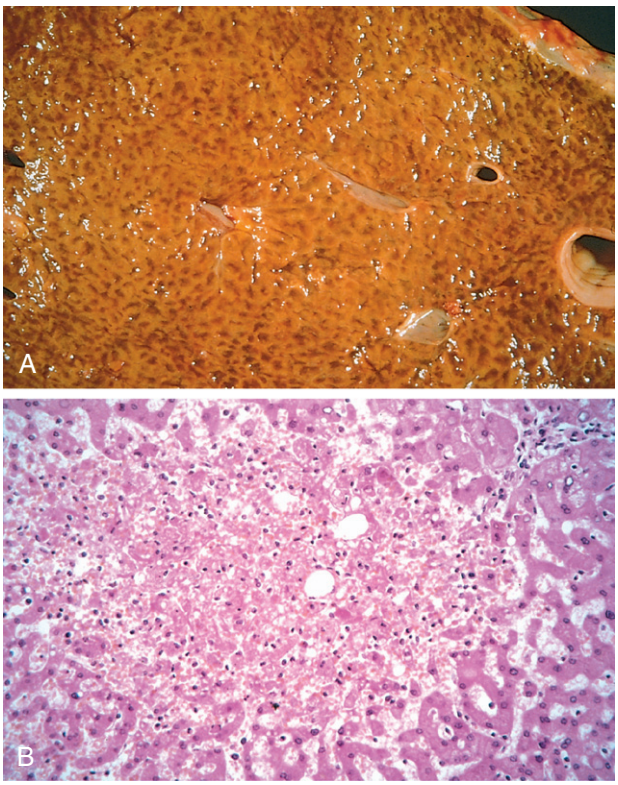


Figure 3–1 Liver with chronic passive congestion and hemorrhagic necrosis. **A,** In this autopsy specimen, central areas are red and slightly depressed compared with the surrounding tan viable parenchyma, creating "nutmeg liver" (so called because it resembles the cut surface of a nutmeg). **B,** Microscopic preparation shows centrilobular hepatic necrosis with hemorrhage and scattered inflammatory cells.
(Courtesy of Dr. James Crawford.)

necrosis, hemorrhage, and hemosiderin-laden macrophages (Fig. 3–1, *B*). In long-standing, severe hepatic congestion (most commonly associated with heart failure), hepatic fibrosis ("cardiac cirrhosis") can develop. Because the central portion of the hepatic lobule is the last to receive blood, centrilobular necrosis also can occur in any setting of reduced hepatic blood flow (including shock from any cause); there need not be previous hepatic congestion.

EDEMA

Approximately 60% of lean body weight is water, two thirds of which is intracellular. Most of the remaining water is found in extracellular compartments in the form of interstitial fluid; only 5% of the body's water is in blood plasma. As noted earlier, *edema* is an accumulation of interstitial fluid within tissues. Extravascular fluid can also collect in body cavities such as the pleural cavity (*hydrothorax*), the pericardial cavity (*hydropericardium*), or the peritoneal cavity (*hydroperitoneum*, or *ascites*). *Anasarca* is severe, generalized edema marked by profound swelling of subcutaneous tissues and accumulation of fluid in body cavities.

Table 3–1 Pathophysiologic Causes of Edema

Increased Hydrostatic Pressure
Impaired Venous Return
Congestive heart failure
Constrictive pericarditis
Ascites (liver cirrhosis)
Venous obstruction or compression
Thrombosis
External pressure (e.g., mass)
Lower extremity inactivity with prolonged dependency
Arteriolar Dilation
Heat
Neurohumoral dysregulation
Reduced Plasma Osmotic Pressure (Hypoproteinemia)
Protein-losing glomerulopathies (nephrotic syndrome)
Liver cirrhosis (ascites)
Malnutrition
Protein-losing gastroenteropathy
Lymphatic Obstruction
Inflammatory
Neoplastic
Postsurgical
Postirradiation
Sodium Retention
Excessive salt intake with renal insufficiency
Increased tubular reabsorption of sodium
Renal hypoperfusion
Increased renin-angiotensin-aldosterone secretion
Inflammation
Acute inflammation
Chronic inflammation
Angiogenesis

Data from Leaf A, Cotran RS: *Renal Pathophysiology*, 3rd ed. New York, Oxford University Press, 1985, p 146.

Table 3–1 lists the major causes of edema. The mechanisms of inflammatory edema are largely related to increased vascular permeability and are discussed in Chapter 2; the *noninflammatory causes* are detailed in the following discussion.

Fluid movement between the vascular and interstitial spaces is governed mainly by two opposing forces—the *vascular hydrostatic pressure* and the *colloid osmotic pressure* produced by plasma proteins. Normally, the outflow of fluid produced by hydrostatic pressure at the arteriolar end of the microcirculation is neatly balanced by inflow due to the slightly elevated osmotic pressure at the venular end; hence there is only a small net outflow of fluid into the interstitial space, which is drained by lymphatic vessels. Either increased hydrostatic pressure or diminished colloid osmotic pressure causes increased movement of water into the interstitium (Fig. 3–2). This in turn increases the tissue hydrostatic pressure, and eventually a new equilibrium is achieved. Excess edema fluid is removed by lymphatic drainage and returned to the bloodstream by way of the thoracic duct (Fig. 3–2).

The edema fluid that accumulates owing to increased hydrostatic pressure or reduced intravascular colloid typically is a protein-poor *transudate*; it has a specific gravity less than 1.012. By contrast, because of increased vascular permeability, inflammatory edema fluid is a protein-rich

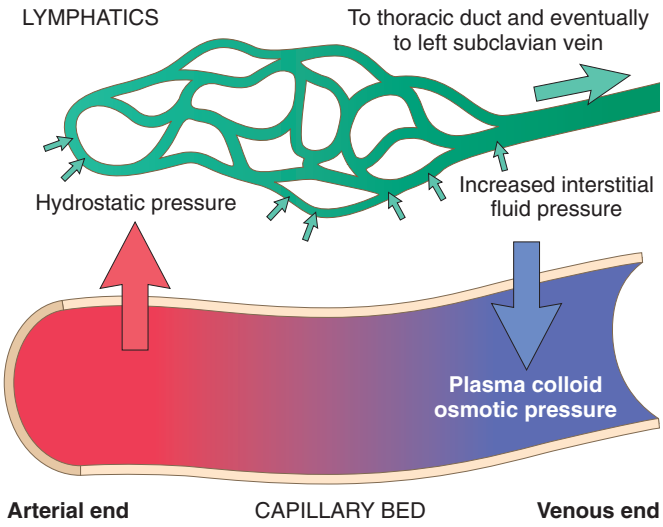


Figure 3–2 Factors influencing fluid movement across capillary walls. Capillary hydrostatic and osmotic forces are normally balanced so there is little net movement of fluid into the interstitium. However, *increased* hydrostatic pressure or *diminished* plasma osmotic pressure leads to extravascular fluid accumulation (edema). Tissue lymphatics drain much of the excess fluid back to the circulation by way of the thoracic duct; however, if the capacity for lymphatic drainage is exceeded, tissue edema results.

exudate with a specific gravity usually greater than 1.020 (see Chapter 2). We will now discuss the various causes of edema.

Increased Hydrostatic Pressure

Local increases in intravascular pressure can result from impaired venous return—for example, a deep venous thrombosis in the lower extremity can cause edema restricted to the distal portion of the affected leg. *Generalized* increases in venous pressure, with resultant systemic edema, occur most commonly in *congestive heart failure* (Chapter 10). Several factors increase venous hydrostatic pressure in patients with congestive heart failure (Fig. 3–3). The reduced cardiac output leads to hypoperfusion of the kidneys, triggering the renin-angiotensin-aldosterone axis and inducing sodium and water retention (*secondary hyperaldosteronism*). In patients with normal heart function, this adaptation increases cardiac filling and cardiac output, thereby improving renal perfusion. However, the failing heart often cannot increase its cardiac output in response to the compensatory increases in blood volume. Instead, a vicious circle of fluid retention, increased venous hydrostatic pressures, and worsening edema ensues. Unless cardiac output is restored or renal water retention is reduced (e.g., by salt restriction or treatment with diuretics or aldosterone antagonists) this downward spiral continues. Because secondary hyperaldosteronism is a common feature of generalized edema, salt restriction, diuretics, and aldosterone antagonists also are of value in the management of generalized edema resulting from other causes.

Reduced Plasma Osmotic Pressure

Under normal circumstances albumin accounts for almost half of the total plasma protein. Therefore conditions in which albumin is either lost from the circulation or synthesized in inadequate amounts are common causes of reduced plasma osmotic pressure. In *nephrotic syndrome* (Chapter 13), damaged glomerular capillaries become leaky, leading to the loss of albumin (and other plasma proteins) in the urine and the development of generalized edema. Reduced albumin synthesis occurs in the setting of severe liver disease (e.g., *cirrhosis*) (Chapter 15) and protein malnutrition (Chapter 7). Regardless of cause, low albumin levels lead in a stepwise fashion to edema, reduced intravascular volume, renal hypoperfusion, and secondary hyperaldosteronism. Unfortunately, increased salt and water retention by the kidney not only fails to correct the plasma volume deficit but also exacerbates the edema, since the primary defect—low serum protein—persists.

Lymphatic Obstruction

Impaired lymphatic drainage and consequent *lymphedema* usually result from a localized obstruction caused by an inflammatory or neoplastic condition. For example, the parasitic infection *filariasis* can cause massive edema of the lower extremity and external genitalia (so-called *elephantiasis*) by engendering inguinal lymphatic and lymph node fibrosis. Infiltration and obstruction of superficial lymphatics by breast cancer may cause edema of the overlying skin; the characteristic finely pitted appearance of the skin of the affected breast is called *peau d'orange* (orange peel). Lymphedema also may occur as a complication of therapy. One relatively common setting for this clinical entity is in women with breast cancer who undergo axillary lymph node resection and/or irradiation, both of which can disrupt and obstruct lymphatic drainage, resulting in severe lymphedema of the arm.

Sodium and Water Retention

Excessive retention of salt (and its obligate associated water) can lead to edema by increasing hydrostatic pressure (due to expansion of the intravascular volume) and

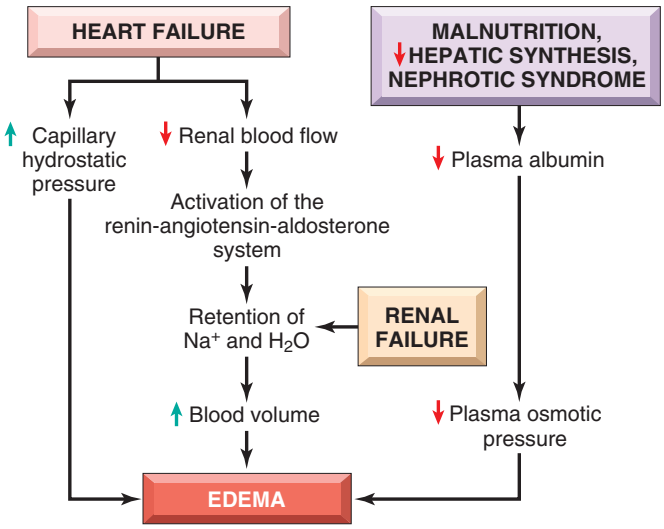


Figure 3–3 Pathways leading to systemic edema due to heart failure, renal failure, or reduced plasma osmotic pressure.

reducing plasma osmotic pressure. Excessive salt and water retention are seen in a wide variety of diseases that compromise renal function, including *poststreptococcal glomerulonephritis* and *acute renal failure* (Chapter 13).

MORPHOLOGY

Edema is easily recognized on gross inspection; microscopic examination shows clearing and separation of the extracellular matrix elements. Although any tissue can be involved, edema most commonly is encountered in subcutaneous tissues, lungs, and brain.

Subcutaneous edema can be diffuse but usually accumulates preferentially in parts of the body positioned the greatest distance below the heart where hydrostatic pressures are highest. Thus, edema typically is most pronounced in the legs with standing and the sacrum with recumbency, a relationship termed **dependent edema.** Finger pressure over edematous subcutaneous tissue displaces the interstitial fluid, leaving a finger-shaped depression; this appearance is called **pitting edema.** Edema due to **renal dysfunction** or **nephrotic syndrome** often manifests first in loose connective tissues (e.g., the eyelids, causing **periorbital edema**). With **pulmonary edema**, the lungs often are two to three times their normal weight, and sectioning reveals frothy, sometimes blood-tinged fluid consisting of a mixture of air, edema fluid, and extravasated red cells. **Brain edema** can be localized (e.g., due to abscess or tumor) or generalized, depending on the nature and extent of the pathologic process or injury. With generalized edema, the sulci are narrowed while the gyri are swollen and flattened against the skull.

Clinical Correlation

The effects of edema vary, ranging from merely annoying to rapidly fatal. Subcutaneous edema is important to recognize primarily because it signals potential underlying cardiac or renal disease; however, when significant, it also can impair wound healing or the clearance of infections. Pulmonary edema is a common clinical problem that most frequently is seen in the setting of left ventricular failure but also may occur in renal failure, acute respiratory distress syndrome (Chapter 11), and inflammatory and infectious disorders of the lung. It can cause death by interfering with normal ventilatory function; besides impeding oxygen diffusion, alveolar edema fluid also creates a favorable environment for infections. Brain edema is life-threatening; if the swelling is severe, the brain can *herniate* (extrude) through the foramen magnum. With increased intracranial pressure, the brain stem vascular supply can be compressed. Either condition can cause death by injuring the medullary centers (Chapter 22).

SUMMARY

Edema

- Edema is the result of the movement of fluid from the vasculature into the interstitial spaces; the fluid may be protein-poor (*transudate*) or protein-rich (*exudate*).
- Edema may be caused by:
  - increased hydrostatic pressure (e.g., heart failure)
  - increased vascular permeability (e.g., inflammation)
  - decreased colloid osmotic pressure, due to reduced plasma albumin
    - decreased synthesis (e.g., liver disease, protein malnutrition)
    - increased loss (e.g., nephrotic syndrome)
  - lymphatic obstruction (e.g., inflammation or neoplasia).
  - sodium retention (e.g., renal failure)

HEMORRHAGE

Hemorrhage, defined as the extravasation of blood from vessels, occurs in a variety of settings. As described earlier, capillary bleeding can occur in chronically congested tissues. The risk of hemorrhage (often after a seemingly insignificant injury) is increased in a wide variety of clinical disorders collectively called *hemorrhagic diatheses.* Trauma, atherosclerosis, or inflammatory or neoplastic erosion of a vessel wall also may lead to hemorrhage, which may be extensive if the affected vessel is a large vein or artery.

Hemorrhage may be manifested by different appearances and clinical consequences.

- Hemorrhage may be external or accumulate within a tissue as a *hematoma,* which ranges in significance from trivial (e.g., a bruise) to fatal (e.g., a massive retroperitoneal hematoma resulting from rupture of a dissecting aortic aneurysm) (Chapter 9).

  Large bleeds into body cavities are given various names according to location—*hemothorax, hemopericardium, hemoperitoneum,* or *hemarthrosis* (in joints). Extensive hemorrhages can occasionally result in jaundice from the massive breakdown of red cells and hemoglobin.
- *Petechiae* are minute (1 to 2 mm in diameter) hemorrhages into skin, mucous membranes, or serosal surfaces (Fig. 3–4, *A*); causes include low platelet counts *(thrombocytopenia),* defective platelet function, and loss of vascular wall support, as in vitamin C deficiency (Chapter 7).
- *Purpura* are slightly larger (3 to 5 mm) hemorrhages. Purpura can result from the same disorders that cause petechiae, as well as trauma, vascular inflammation (*vasculitis*), and increased vascular fragility.
- *Ecchymoses* are larger (1 to 2 cm) subcutaneous hematomas (colloquially called *bruises*). Extravasated red cells are phagocytosed and degraded by macrophages; the characteristic color changes of a bruise are due to the enzymatic conversion of hemoglobin (red-blue color) to bilirubin (blue-green color) and eventually hemosiderin (golden-brown).

The clinical significance of any particular hemorrhage depends on the volume of blood lost and the rate of bleeding. Rapid loss of up to 20% of the blood volume, or slow losses of even larger amounts, may have little impact in healthy adults; greater losses, however, can cause *hemorrhagic (hypovolemic) shock* (discussed later). The site of hemorrhage also is important; bleeding that would be trivial in

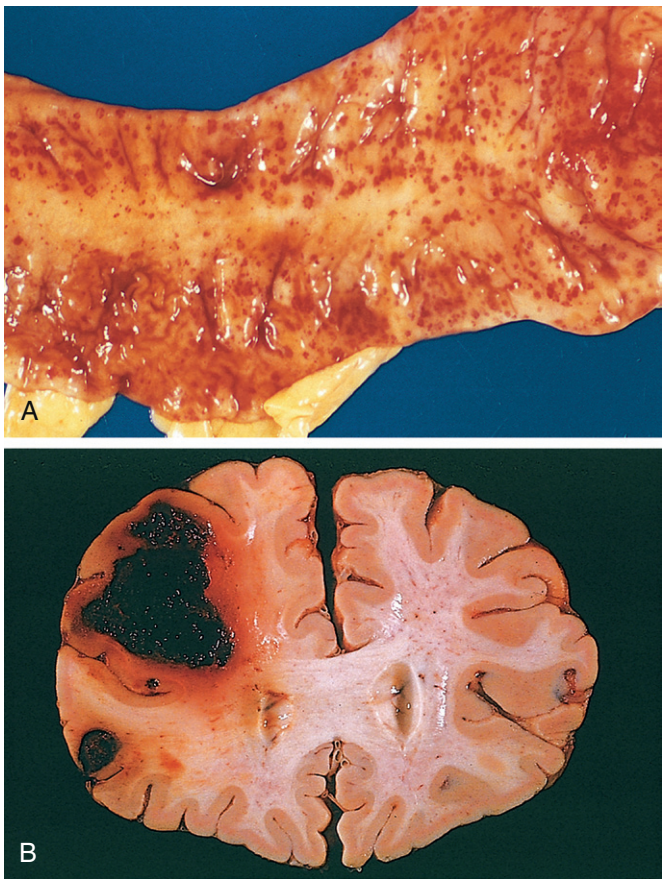


Figure 3–4 A, Punctate petechial hemorrhages of the colonic mucosa, a consequence of thrombocytopenia. **B,** Fatal intracerebral hemorrhage.

the subcutaneous tissues can cause death if located in the brain (Fig. 3–4, *B*). Finally, chronic or recurrent external blood loss (e.g., due to peptic ulcer or menstrual bleeding) frequently culminates in iron deficiency anemia as a consequence of loss of iron in hemoglobin. By contrast, iron is efficiently recycled from phagocytosed red cells, so internal bleeding (e.g., a hematoma) does not lead to iron deficiency.

HEMOSTASIS AND THROMBOSIS

Normal hemostasis comprises a series of regulated processes that maintain blood in a fluid, clot-free state in normal vessels while rapidly forming a localized *hemostatic plug* at the site of vascular injury. The pathologic counterpart of hemostasis is *thrombosis,* the formation of blood clot (*thrombus*) within intact vessels. Both hemostasis and thrombosis involve three elements: the *vascular wall, platelets,* and the *coagulation cascade.* The discussion here begins with normal hemostasis and its regulation.

Normal Hemostasis

The main steps in the process of hemostasis and its regulation are summarized below and shown in Figure 3–5.

- Vascular injury causes transient *arteriolar vasoconstriction* through reflex neurogenic mechanisms, augmented by local secretion of *endothelin* (a potent endothelium-derived vasoconstrictor) (Fig. 3–5, *A*). This effect is fleeting, however, and bleeding would quickly resume if not for the activation of platelets and coagulation factors.
- *Endothelial injury* exposes highly thrombogenic subendothelial extracellular matrix (ECM), facilitating *platelet adherence, activation, and aggregation*. The formation of the initial platelet plug is called *primary hemostasis* (Fig. 3–5, *B*).
- Endothelial injury also exposes *tissue factor* (also known as *factor III* or *thromboplastin*), a membrane-bound procoagulant glycoprotein synthesized by endothelial cells. Exposed tissue factor, acting in conjunction with factor VII (see later), is the major in vivo trigger of the coagulation cascade and its activation eventually culminates in the *activation of thrombin*, which has several roles in regulating coagulation.
- *Activated thrombin* promotes the formation of an insoluble *fibrin* clot by cleaving fibrinogen; thrombin also is a potent activator of additional platelets, which serve to reinforce the hemostatic plug. This sequence, termed secondary hemostasis, results in the formation of a stable clot capable of preventing further hemorrhage (Fig. 3–5, *C*).
- As bleeding is controlled, counterregulatory mechanisms (e.g., factors that produce fibrinolysis, such as *tissue-type plasminogen activator*) are set into motion to ensure that clot formation is limited to the site of injury (Fig. 3–5, *D*).

Discussed next in greater detail are the roles of endothelium, platelets, and the coagulation cascade.

Endothelium

Endothelial cells are central regulators of hemostasis; the balance between the anti- and prothrombotic activities of endothelium determines whether thrombus formation, propagation, or dissolution occurs. Normal endothelial cells express a variety of *anticoagulant* factors that inhibit platelet aggregation and coagulation and promote fibrinolysis; after injury or activation, however, this balance shifts, and endothelial cells acquire numerous *procoagulant* activities (Fig. 3–6). Besides trauma, endothelium can be activated by microbial pathogens, hemodynamic forces, and a number of pro-inflammatory mediators (Chapter 2).

Antithrombotic Properties of Normal Endothelium

Inhibitory Effects on Platelets. Intact endothelium prevents platelets (and plasma coagulation factors) from engaging the highly thrombogenic subendothelial ECM. Nonactivated platelets do not adhere to normal endothelium; even with activated platelets, prostacyclin (i.e., prostaglandin I_2 [PGI_2]) and nitric oxide produced by endothelium impede their adhesion. Both mediators also are potent vasodilators and inhibitors of platelet aggregation; their synthesis by endothelial cells is stimulated by a number of factors (e.g., thrombin, cytokines) produced during coagulation. Endothelial cells also produce adenosine diphosphatase, which degrades adenosine diphosphate (ADP) and further inhibits platelet aggregation (see later).

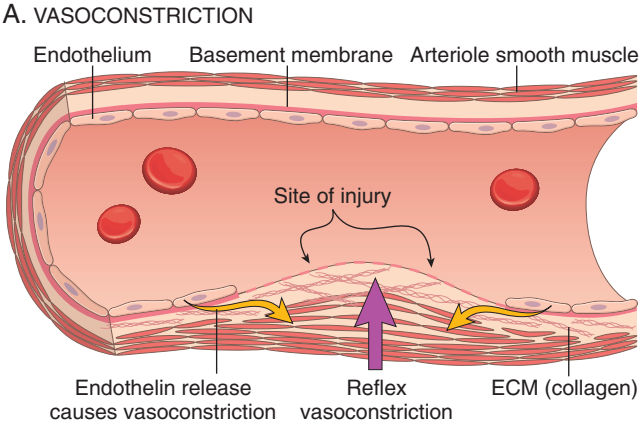


Figure 3–5 Normal hemostasis. **A,** After vascular injury, local neurohumoral factors induce a transient vasoconstriction. **B,** Platelets bind via glycoprotein 1b (Gplb) receptors to von Willebrand factor (vWF) on exposed extracellular matrix (ECM) and are activated, undergoing a shape change and granule release. Released adenosine diphosphate (ADP) and thromboxane A_2 (TxA_2) induce additional platelet aggregation through binding of platelet GpIIb-IIIa receptors to fibrinogen. This platelet aggregate fills the vascular defect, forming the *primary hemostatic plug.* **C,** Local activation of the coagulation cascade (involving tissue factor and platelet phospholipids) results in fibrin polymerization, "cementing" the platelets into a definitive *secondary hemostatic plug* that is larger and more stable than the primary plug and contains entrapped red cells and leukocytes. **D,** Counterregulatory mechanisms, such as release of t-PA (tissue plasminogen activator, a fibrinolytic product) and thrombomodulin (interfering with the coagulation cascade), limit the hemostatic process to the site of injury.

Inhibitory Effects on Coagulation Factors. These actions are mediated by factors expressed on endothelial surfaces, particularly heparin-like molecules, thrombomodulin, and tissue factor pathway inhibitor (Fig. 3–6). *The heparin-like molecules* act indirectly: They are cofactors that greatly enhance the inactivation of thrombin (and other coagulation factors) by the plasma protein *antithrombin III. Thrombomodulin* also acts indirectly: It binds to thrombin, thereby modifying the substrate specificity of thrombin, so that instead of cleaving fibrinogen, it instead cleaves and activates protein C, an anticoagulant. Activated protein C inhibits clotting by cleaving and inactivating two procoagulants, factor Va and factor VIIIa; it requires a cofactor, protein S, which is also synthesized by endothelial cells. Finally, tissue factor pathway inhibitor (TFPI) directly inhibits tissue factor–factor VIIa complex and factor Xa.

Fibrinolysis. Endothelial cells synthesize *tissue-type plasminogen activator*, a protease that cleaves plasminogen to plasmin; plasmin, in turn, cleaves fibrin to degrade thrombi.

Prothrombotic Properties of Injured or Activated Endothelium

Activation of Platelets. Endothelial injury brings platelets into contact with the subendothelial ECM, which includes among its constituents *von Willebrand factor (vWF)*, a large multimeric protein that is synthesized by EC. vWF is held fast to the ECM through interactions with collagen and also binds tightly to Gp1b, a glycoprotein found on the surface of platelets. These interactions allow vWF to act as a sort of molecular glue that binds platelets tightly to denuded vessel walls (Fig. 3–7).

Activation of Clotting Factors. In response to cytokines (e.g., tumor necrosis factor [TNF] or interleukin-1 [IL-1]) or certain bacterial products including endotoxin, endothelial cells produce *tissue factor,* the major in vivo activator of coagulation, and downregulate the expression of thrombomodulin. Activated endothelial cells also bind coagulation factors IXa and Xa (see further on), which augments the catalytic activities of these factors.

Antifibrinolytic Effects. Activated endothelial cells secrete *plasminogen activator inhibitors (PAIs),* which limit fibrinolysis and thereby favor thrombosis.

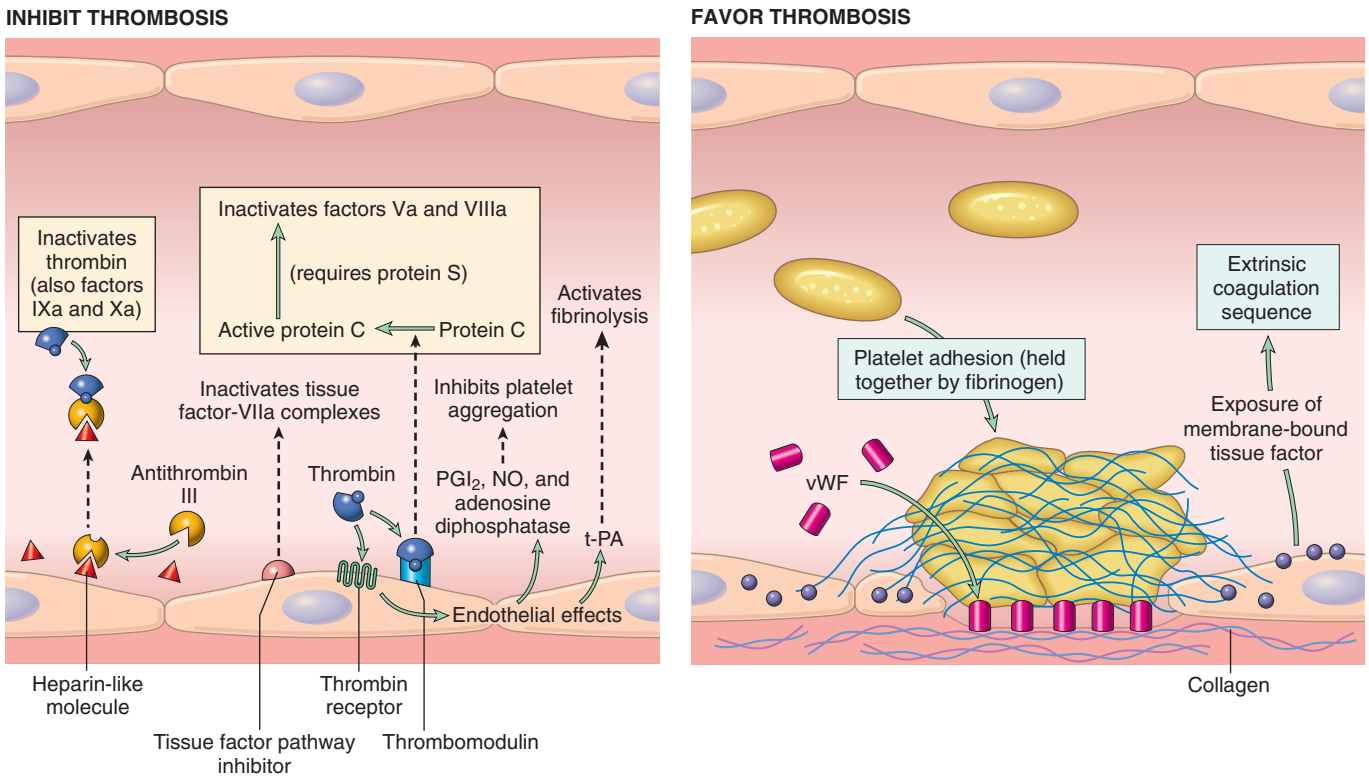


Figure 3–6 Anticoagulant properties of normal endothelium *(left)* and procoagulant properties of injured or activated endothelium *(right)*. NO, nitric oxide; PGI_2, prostaglandin I_2 (prostacyclin); t-PA, tissue plasminogen activator; vWF, von Willebrand factor. Thrombin receptors are also called protease-activated receptors (PARs).

SUMMARY

Endothelial Cells and Coagulation

- Intact, normal endothelial cells help to maintain blood flow by inhibiting the activation of platelets and coagulation factors.
- Endothelial cells stimulated by injury or inflammatory cytokines upregulate expression of procoagulant factors (e.g., tissue factor) that promote clotting, and downregulate expression of anticoagulant factors.
- Loss of endothelial integrity exposes subendothelial vWF and basement membrane collagen, stimulating platelet adhesion, platelet activation, and clot formation.

Platelets

Platelets are anucleate cell fragments shed into the bloodstream by marrow megakaryocytes. They play a critical role in normal hemostasis by forming a hemostatic plug that seals vascular defects, and by providing a surface that recruits and concentrates activated coagulation factors. Platelet function depends on several integrin family glycoprotein receptors, a contractile cytoskeleton, and two types of cytoplasmic granules:

- α *granules,* which express the adhesion molecule P-selectin on their membranes (Chapter 2) and contain

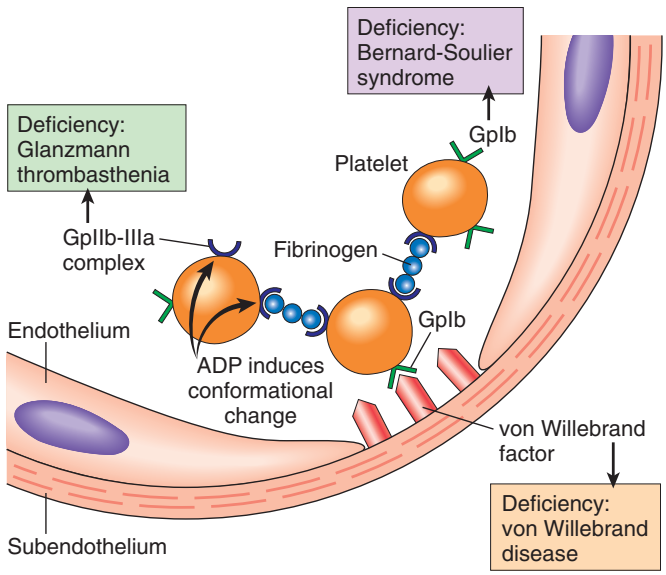


Figure 3–7 Platelet adhesion and aggregation. Von Willebrand factor functions as an *adhesion* bridge between subendothelial collagen and the glycoprotein Ib (GpIb) platelet receptor. Platelet *aggregation* is accomplished by fibrinogen binding to platelet GpIIb-IIIa receptors on different platelets. Congenital deficiencies in the various receptors or bridging molecules lead to the diseases indicated in the *colored boxes.* ADP, adenosine diphosphate.

fibrinogen, fibronectin, factors V and VIII, platelet factor-4 (a heparin-binding chemokine), platelet-derived growth factor (PDGF), and transforming growth factor-β (TGF-β)
- *Dense bodies* (δ granules), which contain adenine nucleotides (ADP and ATP), ionized calcium, histamine, serotonin, and epinephrine

After vascular injury, platelets encounter ECM constituents (collagen is most important) and adhesive glycoproteins such as vWF. This sets in motion a series of events that lead to (1) platelet adhesion, (2) platelet activation, and (3) platelet aggregation (Fig. 3–5, *B*).

Platelet Adhesion

Platelet adhesion initiates clot formation and depends on vWF and platelet glycoprotein Gp1b. Under shear stress (e.g., in flowing blood), vWF undergoes a conformational change, assuming an extended shape that allows it to bind simultaneously to collagen in the ECM and to platelet Gp1b (Fig. 3–7). The importance of this adhesive interaction is highlighted by genetic deficiencies of vWF and Gp1b, both of which result in bleeding disorders—von Willebrand disease (Chapter 11) and Bernard-Soulier disease (a rare condition), respectively.

Platelet Activation

Platelet adhesion leads to an irreversible shape change and secretion (release reaction) of both granule types—a process termed *platelet activation.* Calcium and ADP released from δ granules are especially important in subsequent events, since calcium is required by several coagulation factors and ADP is a potent activator of resting platelets. Activated platelets also synthesize thromboxane A_2 (TxA_2) (Chapter 2), a prostaglandin that activates additional nearby platelets and that also has an important role in platelet aggregation (described below). During activation, platelets undergo a dramatic change in shape from smooth discs to spheres with numerous long, spiky membrane extensions, as well as more subtle changes in the make-up of their plasma membranes. The shape changes enhance subsequent aggregation and increase the surface area available for interaction with coagulation factors. The subtle membrane changes include an increase in the surface expression of negatively charged *phospholipids,* which provide binding sites for both calcium and coagulation factors, and a conformation change in platelet GpIIb/IIIa that permits it to bind fibrinogen.

Platelet Aggregation

Platelet aggregation follows platelet adhesion and activation, and is stimulated by some of the same factors that induce platelet activation, such as TxA_2. Aggregation is promoted by bridging interactions between fibrinogen and GpIIb/IIIa receptors on adjacent platelets (Fig. 3–7). The importance of this interaction is emphasized by a rare inherited deficiency of GpIIb/IIIa (Glanzmann thrombasthenia), which is associated with bleeding and an inability of platelets to aggregate. Recognition of the central role of GpIIb-IIIa receptors in platelet aggregation has stimulated the development of antithrombotic agents that inhibit GpIIb-IIIa function.

Concurrent activation of the coagulation cascade generates thrombin, which stabilizes the platelet plug through two mechanisms:
- Thrombin activates a platelet surface receptor (protease-activated receptor [PAR]), which in concert with ADP and TxA_2 further enhances platelet aggregation. *Platelet contraction* follows, creating an irreversibly fused mass of platelets that constitutes the definitive *secondary hemostatic plug.*
- Thrombin converts fibrinogen to *fibrin* (discussed shortly) within the vicinity of the plug, cementing the platelet plug in place.

Red cells and leukocytes are also found in hemostatic plugs. Leukocytes adhere to platelets by means of P-selectin and to endothelium by various adhesion molecules (Chapter 2); they contribute to the inflammatory response that accompanies thrombosis. Thrombin also promotes inflammation by stimulating neutrophil and monocyte adhesion (described later) and by generating chemotactic *fibrin split products* during fibrinogen cleavage.

Platelet-Endothelial Interactions

The interplay of platelets and endothelium has a profound impact on clot formation. For example, prostaglandin PGI_2 (synthesized by normal endothelium) is a vasodilator and inhibits platelet aggregation, whereas TxA_2 (synthesized by activated platelets, as discussed above) is a potent vasoconstrictor. The balance between the opposing effects of PGI_2 and TxA_2 varies: In normal vessels, PGI_2 effects dominate and platelet aggregation is prevented, whereas endothelial injury decreases PGI_2 production and promotes platelet aggregation and TxA_2 production. The clinical utility of aspirin (an irreversible cyclooxygenase inhibitor) in lowering the risk of coronary thrombosis resides in its ability to permanently block TxA_2 production by platelets, which have no capacity for protein synthesis. Although endothelial PGI_2 production is also inhibited by aspirin, endothelial cells can resynthesize cyclooxygenase, thereby overcoming the blockade. In a manner similar to that for PGI_2, endothelium-derived nitric oxide also acts as a vasodilator and inhibitor of platelet aggregation (Fig. 3–6).

SUMMARY

Platelet Adhesion, Activation, and Aggregation

- Endothelial injury exposes the underlying basement membrane ECM; platelets adhere to the ECM primarily through binding of platelet GpIb receptors to vWF.
- Adhesion leads to platelet activation, an event associated with secretion of platelet granule contents, including calcium (a cofactor for several coagulation proteins) and ADP (a mediator of further platelet activation); dramatic changes in shape and membrane composition; and activation of GpIIb/IIIa receptors.
- The GpIIb/IIIa receptors on activated platelets form bridging crosslinks with fibrinogen, leading to platelet aggregation.
- Concomitant activation of thrombin promotes fibrin deposition, cementing the platelet plug in place.

Coagulation Cascade

The coagulation cascade constitutes the third arm of the hemostatic system. The pathways are schematically presented in Figure 3–8; only general principles are discussed here.

The coagulation cascade is a successive series of amplifying enzymatic reactions. At each step in the process, a proenzyme is proteolyzed to become an active enzyme, which in turn proteolyzes the next proenzyme in the series, eventually leading to the activation of thrombin and the formation of fibrin. *Thrombin has a key role,* as it acts at numerous points in the cascade (highlighted in Fig. 3–8). Thrombin proteolyzes *fibrinogen* into *fibrin* monomers that polymerize into an insoluble gel; this gel encases platelets and other circulating cells in the definitive secondary hemostatic plug. Fibrin polymers are stabilized by the cross-linking activity of factor XIIIa, which also is activated by thrombin.

Each reaction in the pathway depends on the assembly of a complex composed of an *enzyme* (an activated coagulation factor), a *substrate* (a proenzyme form of the next coagulation factor in the series), and a *cofactor* (a reaction accelerator). These components typically are assembled on a *phospholipid surface* (provided by endothelial cells or platelets) and are held together by interactions that depend on *calcium ions* (explaining why blood clotting is prevented

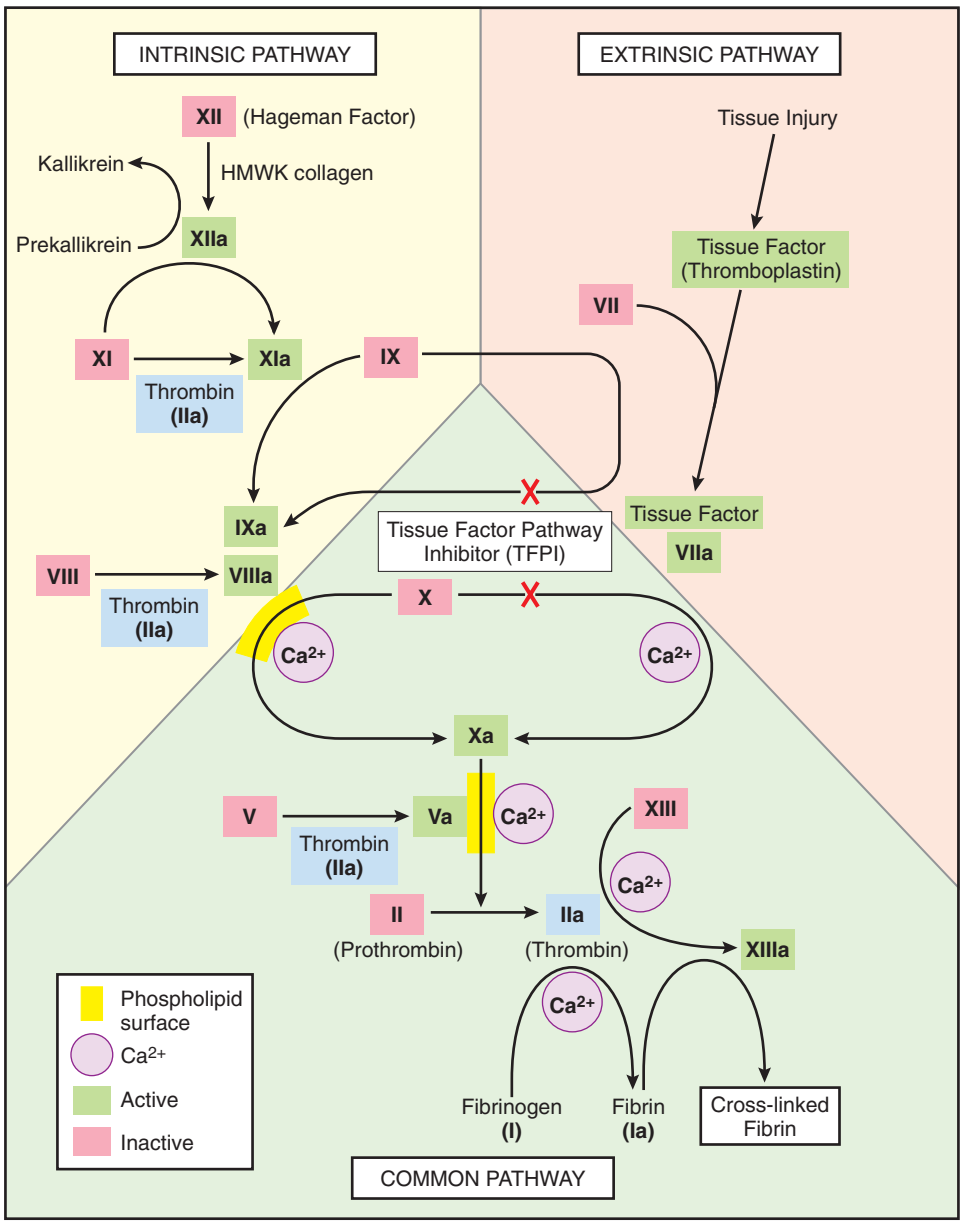


Figure 3–8 The coagulation cascade. Factor IX can be activated by either factor XIa or factor VIIa: In laboratory tests, activation is predominantly dependent on factor XIa, whereas in vivo, factor VIIa appears to be the predominant activator of factor IX. Factors in *red boxes* represent inactive molecules; activated factors, indicated with a lowercase *a*, are in *green boxes*. Note that thrombin (factor IIa) (in *light blue boxes*) contributes to coagulation through multiple positive feedback loops. The *red X's* denote points at which tissue factor pathway inhibitor (TFPI) inhibits activation of factor X and factor IX by factor VIIa. HMWK, high-molecular-weight kininogen; PL, phospholipid.

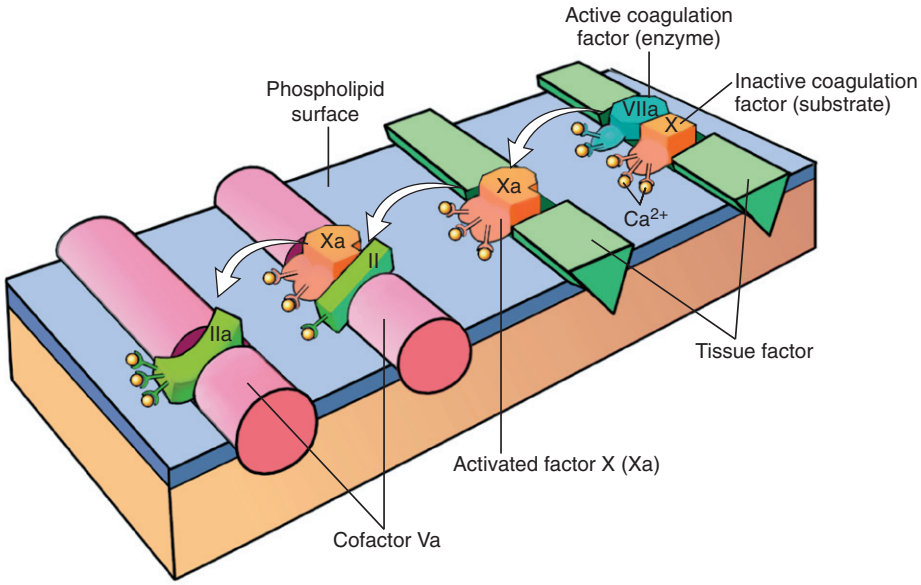


Figure 3–9 Sequential conversion of factor X to factor Xa by way of the extrinsic pathway, followed by conversion of factor II (prothrombin) to factor IIa (thrombin). The initial reaction complex consists of a protease (factor VIIa), a substrate (factor X), and a reaction accelerator (tissue factor) assembled on a platelet phospholipid surface. Calcium ions hold the assembled components together and are essential for the reaction. Activated factor Xa then becomes the protease component of the next complex in the cascade, converting prothrombin to thrombin (factor IIa) in the presence of a different reaction accelerator, factor Va.

by calcium chelators). As shown in Figure 3–9, the sequential cascade of activation can be likened to a "dance" of complexes, with coagulation factors being passed successively from one partner to the next. Parenthetically, the ability of coagulation factors II, VII, IX, and X to bind to calcium requires that additional γ-carboxyl groups be enzymatically appended to certain glutamic acid residues on these proteins. This reaction requires vitamin K as a cofactor and is antagonized by drugs such as *coumadin,* which is widely used as an anticoagulant.

Blood coagulation traditionally is divided into *extrinsic* and *intrinsic* pathways, converging at the activation of factor X (Fig. 3–8). The extrinsic pathway was so designated because it required the addition of an exogenous trigger (originally provided by tissue extracts); the intrinsic pathway only required exposing factor XII (Hageman factor) to a negatively charged surface (even glass suffices). However, this division is largely an artifact of in vitro testing; there are, in fact, several interconnections between the two pathways. The extrinsic pathway is the most physiologically relevant pathway for coagulation occurring after vascular damage; it is activated by *tissue factor,* a membrane-bound glycoprotein expressed at sites of injury.

Clinical labs assess the function of the two arms of the pathway using two standard assays.

- *Prothrombin time* (PT) screens for the activity of the proteins in the extrinsic pathway (factors VII, X, II, V, and fibrinogen). The PT is performed by adding phospholipids and tissue factor to a patient's citrated plasma (sodium citrate chelates calcium and prevents spontaneous clotting), followed by calcium, and the time to fibrin clot formation (usually 11 to 13 seconds) is recorded. Because factor VII is the vitamin K–dependent coagulation factor with the shortest half-life (roughly 7 hours), the PT is used to guide treatment of patients with vitamin K antagonists (e.g., coumadin).
- *Partial thromboplastin time* (PTT) screens for the activity of the proteins in the intrinsic pathway (factors XII, XI, IX, VIII, X, V, II, and fibrinogen). The PTT is performed by adding a negatively charged activator of factor XII (e.g., ground glass) and phospholipids to a patient's citrated plasma, followed by calcium, and recording the time required for clot formation (usually 28 to 35 seconds). The PTT is sensitive to the anticoagulant effects of heparin and is therefore used to monitor its efficacy.

Once thrombin is formed, it not only catalyzes the final steps in the coagulation cascade, but also exerts a wide variety of effects on the local vasculature and inflammatory milieu; it even actively participates in limiting the extent of the hemostatic process (Fig. 3–10). Most of these thrombin-mediated effects occur through *protease-activated receptors (PARs),* which belong to a family of seven-transmembrane-spanning proteins. PARs are present on a variety of cell types, including platelets, endothelium, monocytes, and T lymphocytes. Thrombin activates PARs by clipping their extracellular domains, causing a conformational change that activates associated G proteins. Thus, PAR activation is a catalytic process, explaining the impressive potency of thrombin in eliciting PAR-dependent effects, such as enhancing the adhesive properties of leukocytes.

Once activated, the coagulation cascade must be tightly restricted to the site of injury to prevent inappropriate and potentially dangerous clotting elsewhere in the vascular tree. Besides restricting factor activation to sites of exposed phospholipids, clotting also is controlled by three general categories of natural anticoagulants:

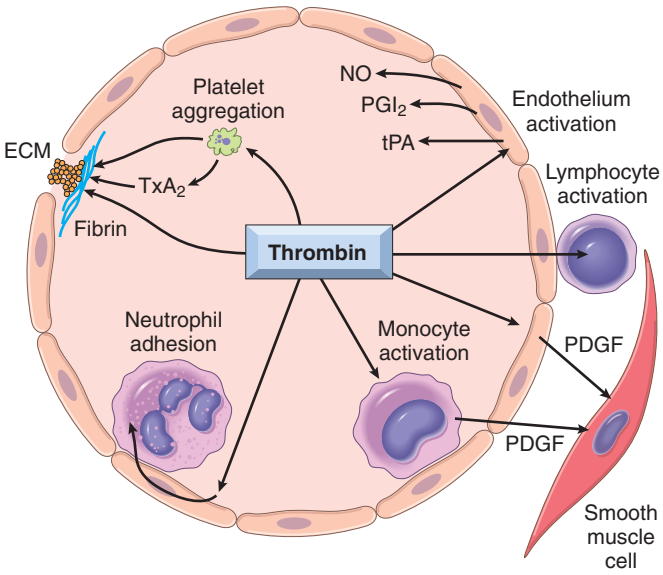


Figure 3–10 Role of thrombin in hemostasis and cellular activation. Thrombin generates fibrin by cleaving fibrinogen, activates factor XIII (which is responsible for cross-linking fibrin into an insoluble clot), and also activates several other coagulation factors, thereby amplifying the coagulation cascade (Fig. 3–8). Through protease-activated receptors (PARs), thrombin activates (1) platelet aggregation and TxA_2 secretion; (2) endothelium, which responds by generating leukocyte adhesion molecules and a variety of fibrinolytic (t-PA), vasoactive (NO, PGI_2), or cytokine (PDGF) mediators; and (3) leukocytes, increasing their adhesion to activated endothelium. ECM, extracellular matrix; NO, nitric oxide; PDGF, platelet-derived growth factor; PGI_2, prostaglandin I_2 (prostacyclin); TxA_2, thromboxane A_2; t-PA, tissue type plasminogen activator. See Figure 3–6 for anticoagulant activities mediated by thrombin via thrombomodulin.

(Courtesy of permission from Shaun Coughlin, MD, PhD, Cardiovascular Research Institute, University of California at San Francisco, San Francisco, California.)

- *Antithrombins* (e.g., antithrombin III) inhibit the activity of thrombin and other serine proteases, namely factors IXa, Xa, XIa, and XIIa. Antithrombin III is activated by binding to heparin-like molecules on endothelial cells—hence the clinical utility of heparin administration to limit thrombosis (Fig. 3–6).
- *Protein C and protein S* are two vitamin K–dependent proteins that act in a complex to proteolytically inactivate cofactors Va and VIIIa. Protein C activation by thrombomodulin was described earlier; protein S is a cofactor for protein C activity (Fig. 3–6).
- *Tissue factor pathway inhibitor (TFPI)* is a protein secreted by endothelium (and other cell types) that inactivates factor Xa and tissue factor–factor VIIa complexes (Fig. 3–8).

Clotting also sets into motion a *fibrinolytic cascade* that moderates the ultimate size of the clot. Fibrinolysis is largely carried out by *plasmin,* which breaks down fibrin and interferes with its polymerization (Fig. 3–11). The resulting *fibrin split products* (*FSPs* or *fibrin degradation products*) also can act as weak anticoagulants. Elevated levels of FSPs (most notably fibrin-derived *D-dimers*) can be used for diagnosing abnormal thrombotic states including disseminated intravascular coagulation (DIC) (Chapter 11), deep venous thrombosis, or pulmonary thromboembolism (described in detail later).

Plasmin is generated by proteolysis of *plasminogen,* an inactive plasma precursor, either by factor XII or by plasminogen activators (Fig. 3–11). The most important of the plasminogen activators is *tissue-type plasminogen activator (t-PA);* t-PA is synthesized principally by endothelial cells and is most active when attached to fibrin. The affinity for fibrin largely confines t-PA fibrinolytic activity to sites of recent thrombosis. *Urokinase-like plasminogen activator* (*u-PA*) is another plasminogen activator present in plasma and in various tissues; it can activate plasmin in the fluid phase. In addition, plasminogen can be cleaved to its active form by the bacterial product *streptokinase,* which is used clinically to lyse clots in some forms of thrombotic disease. As with any potent regulatory component, the activity of plasmin is tightly restricted. To prevent excess plasmin from lysing thrombi indiscriminately throughout the body, free plasmin rapidly complexes with circulating α_2-antiplasmin and is inactivated (Fig. 3–11).

Endothelial cells further modulate the coagulation–anticoagulation balance by releasing *plasminogen activator inhibitors (PAIs);* these block fibrinolysis and confer an overall procoagulation effect (Fig. 3–11). PAI production

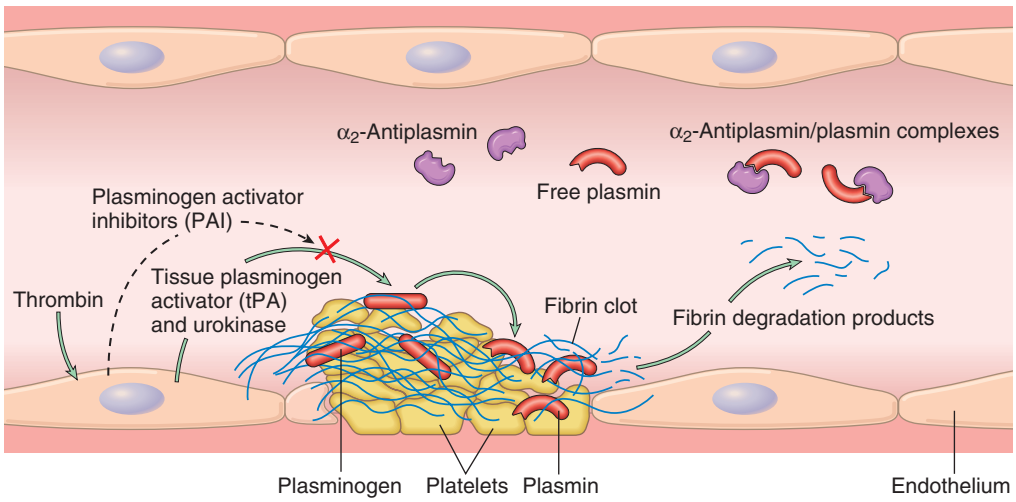


Figure 3–11 The fibrinolytic system, illustrating various plasminogen activators and inhibitors (see text).

is increased by inflammatory cytokines (in particular interferon-γ) and probably contributes to the intravascular thrombosis that accompanies severe inflammation.

SUMMARY

Coagulation Factors

- Coagulation occurs via the sequential enzymatic conversion of a cascade of circulating and locally synthesized proteins.
- Tissue factor elaborated at sites of injury is the most important initiator of the coagulation cascade in vivo.
- At the final stage of coagulation, thrombin converts fibrinogen into insoluble fibrin that contributes to formation of the definitive hemostatic plug.
- Coagulation normally is restricted to sites of vascular injury by
  - limiting enzymatic activation to phospholipid surfaces provided by activated platelets or endothelium
  - natural anticoagulants elaborated at sites of endothelial injury or during activation of the coagulation cascade
  - expression of thrombomodulin on normal endothelial cells, which binds thrombin and converts it into an anticoagulant
  - activation of fibrinolytic pathways (e.g., by association of tissue plasminogen activator with fibrin)

Thrombosis

Having reviewed the process of normal hemostasis, we now turn to the three primary *abnormalities that lead to thrombus formation (called Virchow's triad)*: (1) endothelial injury, (2) stasis or turbulent blood flow, and (3) hypercoagulability of the blood (Fig. 3–12).

Endothelial Injury

Endothelial injury is an important cause of thrombosis, particularly in the heart and the arteries, where high flow rates might otherwise impede clotting by preventing platelet adhesion or diluting coagulation factors. Examples of thrombosis related to endothelial damage are the formation of thrombi in the cardiac chambers after myocardial infarction, over ulcerated plaques in atherosclerotic arteries, or at sites of traumatic or inflammatory vascular injury (*vasculitis*). Overt loss of endothelium exposes subendothelial ECM (leading to platelet adhesion), releases tissue factor, and reduces local production of PGI_2 and plasminogen activators. Of note, however, *endothelium need not be denuded or physically disrupted to contribute to the development of thrombosis; any perturbation in the dynamic balance of the prothrombotic and antithrombotic effects of endothelium can influence clotting locally.* Thus, dysfunctional endothelium elaborates greater amounts of procoagulant factors (e.g., platelet adhesion molecules, tissue factor, PAI) and synthesizes lesser amounts of anticoagulant molecules (e.g., thrombomodulin, PGI_2, t-PA). Endothelial dysfunction can be induced by a variety of insults, including hypertension, turbulent blood flow, bacterial products, radiation injury, metabolic abnormalities such as homocystinuria and hypercholesterolemia, and toxins absorbed from cigarette smoke.

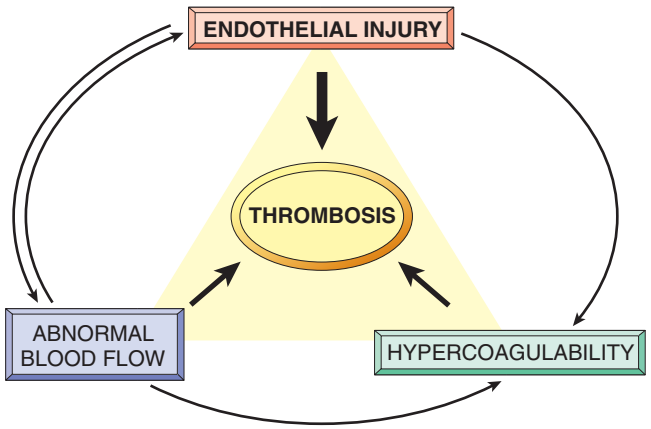


Figure 3–12 Virchow's triad in thrombosis. Endothelial integrity is the most important factor. Abnormalities of procoagulants or anticoagulants can tip the balance in favor of thrombosis. Abnormal blood flow (stasis or turbulence) can lead to hypercoagulability directly and also indirectly through endothelial dysfunction.

Abnormal Blood Flow

Turbulence contributes to arterial and cardiac thrombosis by causing endothelial injury or dysfunction, as well as by forming countercurrents and local pockets of stasis. *Stasis* is a major factor in the development of venous thrombi. Under conditions of normal *laminar* blood flow, platelets (and other blood cells) are found mainly in the center of the vessel lumen, separated from the endothelium by a slower-moving layer of plasma. By contrast, stasis and turbulent (chaotic) blood flow have the following deleterious effects:

- Both promote endothelial cell activation and enhanced procoagulant activity, in part through flow-induced changes in endothelial gene expression.
- Stasis allows platelets and leukocytes to come into contact with the endothelium when the flow is sluggish.
- Stasis also slows the washout of activated clotting factors and impedes the inflow of clotting factor inhibitors.

Turbulent and static blood flow contribute to thrombosis in a number of clinical settings. Ulcerated atherosclerotic plaques not only expose subendothelial ECM but also cause turbulence. Abnormal aortic and arterial dilations called *aneurysms* create local stasis and consequently a fertile site for thrombosis (Chapter 9). Acute myocardial infarction results in focally noncontractile myocardium. Ventricular remodeling after more remote infarction can lead to aneurysm formation. In both cases, cardiac mural thrombi are more easily formed due to the local blood stasis (Chapter 10). Mitral valve stenosis (e.g., after rheumatic heart disease) results in left atrial dilation. In conjunction with atrial fibrillation, a dilated atrium is a site of profound stasis and a prime location for the development of thrombi. *Hyperviscosity syndromes* (such as *polycythemia*) (Chapter 11) increase resistance to flow and cause small vessel stasis; the deformed red cells in sickle cell anemia (Chapter 11) cause vascular occlusions, and the resultant stasis also predisposes to thrombosis.

Hypercoagulability

Hypercoagulability contributes infrequently to arterial or intracardiac thrombosis but is an important underlying risk factor for venous thrombosis. It is loosely defined as any alteration of the coagulation pathways that predisposes affected persons to thrombosis, and can be divided into primary (genetic) and secondary (acquired) disorders (Table 3–2).

Primary (inherited) hypercoagulability most often is caused by mutations in the factor V and prothrombin genes:

- Approximately 2% to 15% of whites carry a specific factor V mutation (called the Leiden mutation, after the Dutch city where it was first described). The mutation alters an amino acid residue in factor V and renders it resistant to protein C. Thus, an important antithrombotic counter-regulatory mechanism is lost. Heterozygotes carry a 5-fold increased risk for venous thrombosis, with homozygotes having a 50-fold increased risk.
- A single-nucleotide substitution (G to A) in the 3′-untranslated region of the prothrombin gene is a fairly common allele (found in 1% to 2% of the general population). This variant results in increased prothrombin transcription and is associated with a nearly threefold increased risk for venous thromboses.

Table 3–2 Hypercoagulable States

Primary (Genetic)
Common (>1% of the Population)
Factor V mutation (G1691A mutation; factor V Leiden)
Prothrombin mutation (G20210A variant)
5,10-Methylene tetrahydrofolate reductase (homozygous C677T mutation)
Increased levels of factor VIII, IX, or XI or fibrinogen
Rare
Antithrombin III deficiency
Protein C deficiency
Protein S deficiency
Very Rare
Fibrinolysis defects
Homozygous homocystinuria (deficiency of cystathione β-synthetase)
Secondary (Acquired)
High Risk for Thrombosis
Prolonged bed rest or immobilization
Myocardial infarction
Atrial fibrillation
Tissue injury (surgery, fracture, burn)
Cancer
Prosthetic cardiac valves
Disseminated intravascular coagulation
Heparin-induced thrombocytopenia
Antiphospholipid antibody syndrome
Lower Risk for Thrombosis
Cardiomyopathy
Nephrotic syndrome
Hyperestrogenic states (pregnancy and postpartum)
Oral contraceptive use
Sickle cell anemia
Smoking

- Less common primary hypercoagulable states include inherited deficiencies of anticoagulants such as antithrombin III, protein C, or protein S; affected patients typically present with venous thrombosis and recurrent thromboembolism in adolescence or early adult life. Congenitally elevated levels of homocysteine contribute to arterial and venous thromboses (and indeed to the development of atherosclerosis) (Chapter 9).

Although the risk of thrombosis is only mildly increased in heterozygous carriers of factor V Leiden and the prothrombin gene variant, these genetic factors carry added significance for two reasons. First, both abnormal alleles are sufficiently frequent that homozygous and compound heterozygous persons are not uncommon, and these individuals are at much higher risk for thrombosis. More importantly, heterozygous individuals are at higher risk for venous thrombosis in the setting of other acquired risk factors, such as pregnancy, prolonged bed rest, and lengthy airplane flights. Consequently, *inherited causes of hypercoagulability should be considered in young patients (<50 years of age), even when other acquired risk factors are present*.

Secondary (acquired) hypercoagulability is seen in many settings (Table 3–2). In some situations (e.g., cardiac failure or trauma), stasis or vascular injury may be the most important factor. The hypercoagulability associated with oral contraceptive use and the hyperestrogenic state of pregnancy may be related to increased hepatic synthesis of coagulation factors and reduced synthesis of antithrombin III. In disseminated cancers, release of procoagulant tumor products (e.g., mucin from adenocarcinoma) predisposes to thrombosis. The hypercoagulability seen with advancing age has been attributed to increased platelet aggregation and reduced release of PGI_2 from endothelium. Smoking and obesity promote hypercoagulability by unknown mechanisms.

Among the acquired thrombophilic states, two are particularly important clinical problems and deserve special mention:

- *Heparin-induced thrombocytopenic (HIT) syndrome.* This syndrome occurs in up to 5% of patients treated with *unfractionated* heparin (for therapeutic anticoagulation). It is marked by the development of autoantibodies that bind complexes of heparin and platelet membrane protein (platelet factor-4) (Chapter 11). Although the mechanism is unclear, it appears that these antibodies may also bind similar complexes present on platelet and endothelial surfaces, resulting in platelet activation, aggregation, and consumption (hence *thrombocytopenia*), as well as causing endothelial cell injury. The overall result is a *prothrombotic state,* even in the face of heparin administration and low platelet counts. Newer low-molecular-weight *fractionated* heparin preparations induce autoantibodies less frequently but can still cause thrombosis if antibodies have already formed.
- *Antiphospholipid antibody syndrome.* This syndrome has protean manifestations, including recurrent thrombosis, repeated miscarriages, cardiac valve vegetations, and thrombocytopenia; it is associated with autoantibodies directed against anionic phospholipids (e.g., cardiolipin) or—more accurately—plasma protein antigens that are unveiled by binding to such phospholipids (e.g., prothrombin). In vivo, these antibodies induce a

hypercoagulable state, perhaps by inducing endothelial injury, by activating platelets or complement directly, or by interacting with the catalytic domains of certain coagulation factors. In vitro (in the absence of platelets and endothelium), however, the antibodies interfere with phospholipid complex assembly, thereby inhibiting coagulation (hence the designation *lupus anticoagulant*). In patients with anticardiolipin antibodies, serologic testing for syphilis will yield a false-positive result, because the antigen in the standard assays is embedded in cardiolipin.

Patients with antiphospholipid antibody syndrome fall into two categories. Many have *secondary antiphospholipid syndrome* due to a well-defined autoimmune disease, such as systemic lupus erythematosus (Chapter 4). The remainder of these patients exhibit only the manifestations of a hypercoagulable state without evidence of another autoimmune disorder (*primary antiphospholipid syndrome*). Although antiphospholipid antibodies are associated with thrombotic diatheses, they also occur in 5% to 15% of apparently normal persons; the implication is that their presence may be necessary but not sufficient to cause full-blown antiphospholipid antibody syndrome.

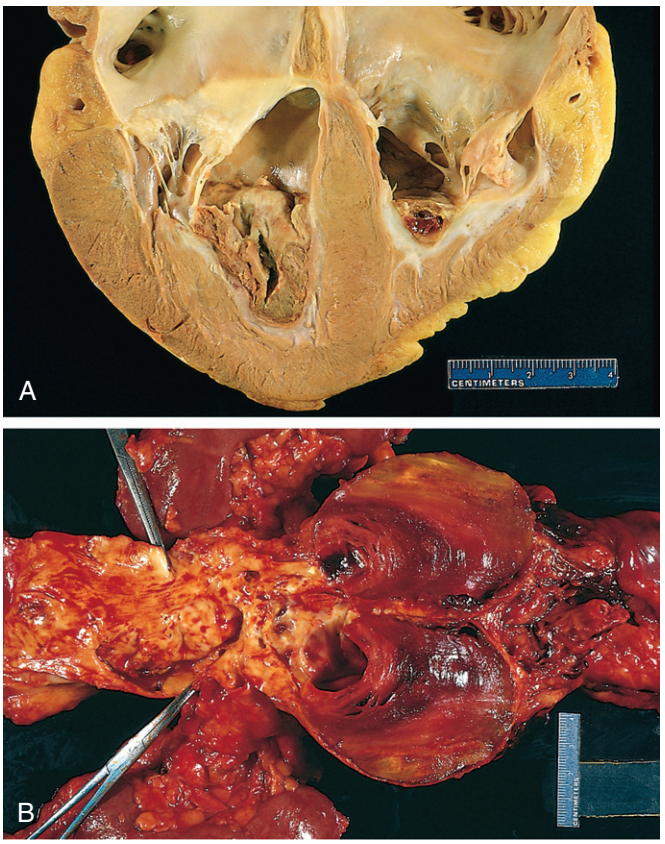


Figure 3–13 Mural thrombi. **A,** Thrombus in the left and right ventricular apices, overlying white fibrous scar. **B,** Laminated thrombus in a dilated abdominal aortic aneurysm. Numerous friable mural thrombi are also superimposed on advanced atherosclerotic lesions of the more proximal aorta (*left side of photograph*).

MORPHOLOGY

Thrombi can develop anywhere in the cardiovascular system. Arterial or cardiac thrombi typically arise at sites of endothelial injury or turbulence; venous thrombi characteristically occur at sites of stasis. Thrombi are focally attached to the underlying vascular surface and tend to propagate **toward** the heart; thus, arterial thrombi grow in a retrograde direction from the point of attachment, while venous thrombi extend in the direction of blood flow. The propagating portion of a thrombus tends to be poorly attached and therefore prone to fragmentation and migration through the blood as an **embolus.**

Thrombi can have grossly (and microscopically) apparent laminations called **lines of Zahn;** these represent pale platelet and fibrin layers alternating with darker red cell–rich layers. Such lines are significant in that they are only found in thrombi that form in flowing blood; their presence can therefore usually distinguish antemortem thrombosis from the bland nonlaminated clots that form in the postmortem state. Although thrombi formed in the "low-flow" venous system superficially resemble postmortem clots, careful evaluation generally reveals ill-defined laminations.

Thrombi occurring in heart chambers or in the aortic lumen are designated **mural thrombi.** Abnormal myocardial contraction (arrhythmias, dilated cardiomyopathy, or myocardial infarction) or endomyocardial injury (myocarditis, catheter trauma) promote cardiac mural thrombi (Fig. 3–13, *A*), while ulcerated atherosclerotic plaques and aneurysmal dilation promote aortic thrombosis (Fig. 3-13, *B*).

Arterial thrombi are typically relatively rich in platelets, as the processes underlying their development (e.g., endothelial injury) lead to platelet activation. Although usually superimposed on a ruptured atherosclerotic plaque, other vascular injuries (vasculitis, trauma) can also be causal. **Venous thrombi (phlebothrombosis)** frequently propagate some distance toward the heart, forming a long cast within the vessel lumen that is prone to give rise to emboli. An increase in the activity of coagulation factors is involved in the genesis of most venous thrombi, with platelet activation playing a secondary role. Because these thrombi form in the sluggish venous circulation, they tend to contain more enmeshed red cells, leading to the moniker **red,** or **stasis, thrombi.** The veins of the lower extremities are most commonly affected (90% of venous thromboses); however, venous thrombi also can occur in the upper extremities, periprostatic plexus, or ovarian and periuterine veins, and under special circumstances may be found in the dural sinuses, portal vein, or hepatic vein.

At autopsy, **postmortem clots** can sometimes be mistaken for venous thrombi. However, the former are gelatinous and due to red cell settling have a dark red dependent portion and a yellow "chicken fat" upper portion; they also are usually not attached to the underlying vessel wall. By contrast, red thrombi typically are firm, focally attached to vessel walls, and contain gray strands of deposited fibrin.

Thrombi on heart valves are called **vegetations.** Bacterial or fungal blood-borne infections can cause valve damage, leading to the development of large thrombotic masses **(infective endocarditis)** (Chapter 10). Sterile vegetations also can develop on noninfected valves in hypercoagulable

states—the lesions of so-called **nonbacterial thrombotic endocarditis** (Chapter 10). Less commonly, sterile, **verrucous endocarditis (Libman-Sacks endocarditis)** can occur in the setting of systemic lupus erythematosus (Chapter 4).

Fate of the Thrombus

If a patient survives an initial thrombotic event, over the ensuing days to weeks the thrombus evolves through some combination of the following four processes:

- *Propagation.* The thrombus enlarges through the accretion of additional platelets and fibrin, increasing the odds of vascular occlusion or embolization.
- *Embolization.* Part or all of the thrombus is dislodged and transported elsewhere in the vasculature.
- *Dissolution.* If a thrombus is newly formed, activation of fibrinolytic factors may lead to its rapid shrinkage and complete dissolution. With older thrombi, extensive fibrin polymerization renders the thrombus substantially more resistant to plasmin-induced proteolysis, and lysis is ineffectual. This acquisition of resistance to lysis has clinical significance, as therapeutic administration of fibrinolytic agents (e.g., t-PA in the setting of acute coronary thrombosis) generally is not effective unless given within a few hours of thrombus formation.
- *Organization and recanalization.* Older thrombi become *organized* by the ingrowth of endothelial cells, smooth muscle cells, and fibroblasts into the fibrin-rich thrombus (Fig. 3–14). In time, capillary channels are formed that—to a limited extent—create conduits along the length of the thrombus, thereby reestablishing the continuity of the original lumen. Further recanalization can sometimes convert a thrombus into a vascularized mass of connective tissue that is eventually incorporated into the wall of the remodeled vessel. Occasionally, instead of organizing, the center of a thrombus undergoes enzymatic digestion, presumably because of the release of lysosomal enzymes from entrapped leukocytes. If bacterial seeding occurs, the contents of degraded thrombi serve as an ideal culture medium, and the resulting infection may weaken the vessel wall, leading to formation of a *mycotic aneurysm* (Chapter 9).

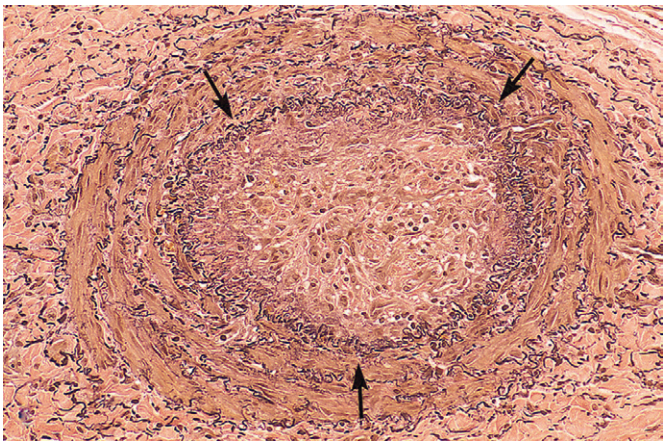


Figure 3–14 Low-power view of a thrombosed artery stained for elastic tissue. The original lumen is delineated by the internal elastic lamina *(arrows)* and is totally filled with organized thrombus.

Clinical Correlation

Thrombi are significant because *they cause obstruction of arteries and veins and may give rise to emboli.* Which effect is of greatest clinical importance depends on the site of thrombosis. Thus, while venous thrombi can cause congestion and edema in vascular beds distal to an obstruction, they are most worrisome because of their potential to embolize to the lungs and cause death. Conversely, while arterial thrombi can embolize and cause tissue infarction, their tendency to obstruct vessels (e.g., in coronary and cerebral vessels) is considerably more important.

Venous Thrombosis (Phlebothrombosis). Most venous thrombi occur in either the superficial or the deep veins of the leg. Superficial venous thrombi usually arise in the saphenous system, particularly in the setting of varicosities; these rarely embolize but can be painful and can cause local congestion and swelling from impaired venous outflow, predisposing the overlying skin to development of infections and *varicose ulcers.* Deep venous thromboses ("DVTs") in the *larger leg veins at or above the knee joint* (e.g., popliteal, femoral, and iliac veins) are more serious because they are prone to embolize. Although such DVTs may cause local pain and edema, the venous obstruction often is circumvented by collateral channels. Consequently, DVTs are entirely asymptomatic in *approximately 50% of patients* and are recognized only after they have embolized to the lungs.

Lower-extremity DVTs are associated with stasis and hypercoagulable states, as described earlier (Table 3–2); thus, common predisposing factors include congestive heart failure, bed rest and immobilization; the latter two factors reduce the milking action of leg muscles and thus slow venous return. Trauma, surgery, and burns not only immobilize a patient but are also associated with vascular injury, procoagulant release, increased hepatic synthesis of coagulation factors, and reduced t-PA production. Many factors contribute to the thrombotic diathesis of pregnancy; besides the potential for amniotic fluid infusion into the circulation at the time of delivery, pressure produced by the enlarging fetus and uterus can produce stasis in the veins of the legs, and late pregnancy and the postpartum period are associated with hypercoagulability. Tumor-associated procoagulant release is largely responsible for the increased risk of thromboembolic phenomena seen in disseminated cancers, which is sometimes referred to as *migratory thrombophlebitis* due to its tendency to transiently involve several different venous beds, or as *Trousseau syndrome,* for Armand Trousseau, who both described the disorder and suffered from it. Regardless of the specific clinical setting, the risk of DVT is increased in persons over age 50.

While the many conditions that predispose to thrombosis are well recognized, the phenomenon remains unpredictable. It occurs at a distressingly high frequency in otherwise healthy and ambulatory people without apparent provocation or underlying abnormality. Equally important is that asymptomatic thrombosis (and presumably subsequent resolution) occurs considerably more frequently than is generally appreciated.

SUMMARY

Thrombosis

- Thrombus development usually is related to one or more components of Virchow's triad:
  - endothelial injury (e.g., by toxins, hypertension, inflammation, or metabolic products)
  - abnormal blood flow, stasis or turbulence (e.g., due to aneurysms, atherosclerotic plaque)
  - hypercoagulability: either primary (e.g., factor V Leiden, increased prothrombin synthesis, antithrombin III deficiency) or secondary (e.g., bed rest, tissue damage, malignancy)
- Thrombi may propagate, resolve, become organized, or embolize.
- Thrombosis causes tissue injury by local vascular occlusion or by distal embolization.

Disseminated Intravascular Coagulation

Disseminated intravascular coagulation (DIC) is the sudden or insidious onset of widespread thrombosis within the microcirculation. It may be seen in disorders ranging from obstetric complications to advanced malignancy. The thrombi are generally microscopic in size, yet so numerous as to often cause circulatory insufficiency, particularly in the brain, lungs, heart, and kidneys. To complicate matters, the widespread microvascular thrombosis consumes platelets and coagulation proteins (hence the synonym *consumption coagulopathy*), and at the same time, fibrinolytic mechanisms are activated. Thus, an initially thrombotic disorder can evolve into a bleeding catastrophe. A point worthy of emphasis is that *DIC is not a primary disease but rather a potential complication of numerous conditions associated with widespread activation of thrombin*. It is discussed in greater detail along with other bleeding diatheses in Chapter 11.

EMBOLISM

An embolus is an intravascular solid, liquid, or gaseous mass that is carried by the blood to a site distant from its point of origin. The vast majority of emboli derive from a dislodged thrombus—hence the term *thromboembolism.* Less common types of emboli include fat droplets, bubbles of air or nitrogen, atherosclerotic debris (*cholesterol emboli*), tumor fragments, bits of bone marrow, and amniotic fluid. Inevitably, emboli lodge in vessels too small to permit further passage, resulting in partial or complete vascular occlusion; depending on the site of origin, emboli can lodge anywhere in the vascular tree. The primary consequence of systemic embolization is ischemic necrosis (*infarction*) of downstream tissues, while embolization in the pulmonary circulation leads to hypoxia, hypotension, and right-sided heart failure.

Pulmonary Thromboembolism

The incidence of pulmonary embolism is 2 to 4 per 1000 hospitalized patients. Although the rate of fatal pulmonary embolus (PE) has declined from 6% to 2% over the last quarter-century, pulmonary embolism still causes about 200,000 deaths per year in the United States. In greater than 95% of cases, venous emboli originate from thrombi within deep leg veins proximal to the popliteal fossa; embolization from lower leg thrombi is uncommon.

Fragmented thrombi from DVTs are carried through progressively larger channels and usually pass through the right side of the heart before arresting in the pulmonary vasculature. Depending on size, a PE can occlude the main pulmonary artery, lodge at the bifurcation of the right and left pulmonary arteries (*saddle embolus*), or pass into the smaller, branching arterioles (Fig. 3–15). Frequently, multiple emboli occur, either sequentially or as a shower of smaller emboli from a single large thrombus; *a patient who has had one pulmonary embolus is at increased risk for having more.* Rarely, an embolus passes through an atrial or ventricular defect and enters the systemic circulation (*paradoxical embolism*). A more complete discussion of PE is found in Chapter 12; the major clinical and pathologic features are the following:

- Most pulmonary emboli (60% to 80%) are small and clinically silent. With time, they undergo organization and become incorporated into the vascular wall; in some cases, organization of thromboemboli leaves behind bridging fibrous *webs.*
- At the other end of the spectrum, a large embolus that blocks a major pulmonary artery can cause sudden death.
- Embolic obstruction of medium-sized arteries and subsequent rupture of capillaries rendered anoxic can cause pulmonary hemorrhage. Such embolization does not usually cause pulmonary infarction since the area also receives blood through an intact bronchial circulation (dual circulation). However, a similar embolus in the setting of left-sided cardiac failure (and diminished bronchial artery perfusion) can lead to a pulmonary infarct.
- Embolism to small end-arteriolar pulmonary branches usually causes infarction.
- Multiple emboli occurring over time can cause pulmonary hypertension and right ventricular failure (cor pulmonale).

Figure 3–15 Embolus derived from a lower-extremity deep venous thrombus lodged in a pulmonary artery branch.

Systemic Thromboembolism

Most systemic emboli (80%) arise from intracardiac mural thrombi; two thirds are associated with left ventricular infarcts and another 25% with dilated left atria (e.g., secondary to mitral valve disease). The remainder originate from aortic aneurysms, thrombi overlying ulcerated atherosclerotic plaques, fragmented valvular vegetations (Chapter 10), or the venous system (*paradoxical emboli*); 10% to 15% of systemic emboli are of unknown origin.

By contrast with venous emboli, which lodge primarily in the lung, arterial emboli can travel virtually anywhere; their final resting place understandably depends on their point of origin and the relative flow rates of blood to the downstream tissues. Common arteriolar *embolization sites* include the lower extremities (75%) and central nervous system (10%); intestines, kidneys, and spleen are less common targets. The consequences of embolization depend on the caliber of the occluded vessel, the collateral supply, and the affected tissue's vulnerability to anoxia; arterial emboli often lodge in end arteries and cause infarction.

Fat Embolism

Soft tissue crush injury or rupture of marrow vascular sinusoids (long bone fracture) releases microscopic fat globules into the circulation. Fat and marrow emboli are common incidental findings after vigorous cardiopulmonary resuscitation but probably are of little clinical consequence. Similarly, although fat and marrow embolism occurs in some 90% of individuals with severe skeletal injuries (Fig. 3–16, *A*), less than 10% show any clinical findings. However, a minority of patients develop a symptomatic *fat embolism syndrome* characterized by *pulmonary insufficiency, neurologic symptoms, anemia, thrombocytopenia, and a diffuse petechial rash, which is fatal in 10% of cases*. Clinical signs and symptoms appear 1 to 3 days after injury as the sudden onset of tachypnea, dyspnea, tachycardia, irritability, and restlessness, which can progress rapidly to delirium or coma.

The pathogenesis of fat emboli syndrome involves both mechanical obstruction and biochemical injury. Fat microemboli occlude pulmonary and cerebral microvasculature, both directly and by triggering platelet aggregation. This deleterious effect is exacerbated by fatty acid release from lipid globules, which causes local toxic endothelial injury. Platelet activation and granulocyte recruitment (with free radical, protease, and eicosanoid release) (Chapter 2) complete the vascular assault. Because lipids are dissolved by the solvents used during tissue-processing, microscopic demonstration of fat microglobules (i.e., in the absence of accompanying marrow elements) requires specialized techniques (frozen sections and fat stains).

Amniotic Fluid Embolism

Amniotic fluid embolism is an uncommon, grave complication of labor and the immediate postpartum period (1 in 40,000 deliveries). The mortality rate approaches 80%, making it the most common cause of maternal death in the developed world; it accounts for 10% of maternal deaths in the United States, while 85% of survivors suffer some form of permanent neurologic deficit. Onset is characterized by sudden severe dyspnea, cyanosis, and hypotensive shock, followed by seizures and coma. If the patient survives the initial crisis, pulmonary edema typically develops, along with (in about half the patients) disseminated intravascular coagulation secondary to release of thrombogenic substances from amniotic fluid.

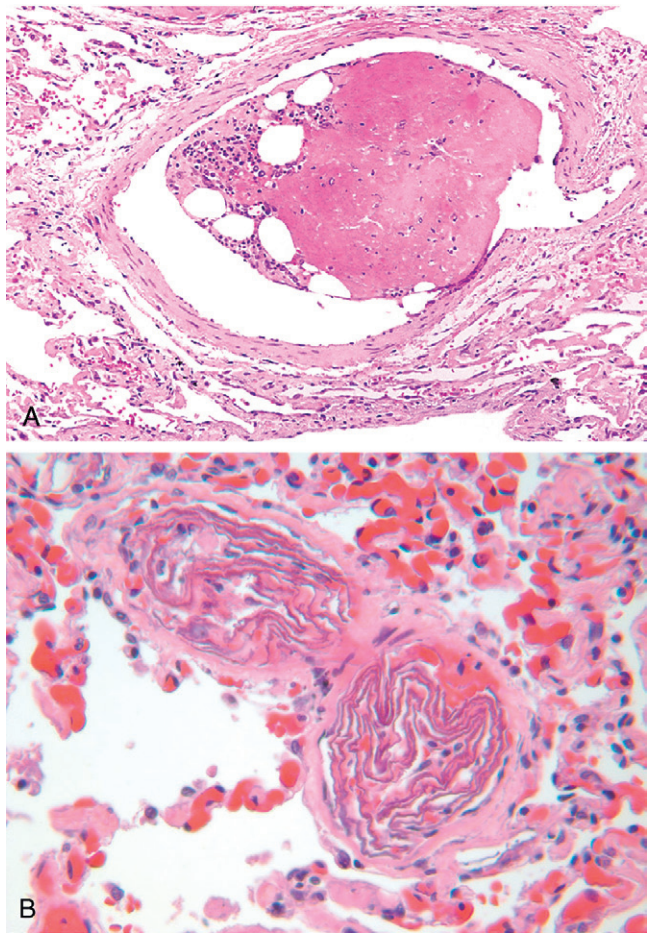


Figure 3–16 Unusual types of emboli. **A,** Bone marrow embolus. The embolus is composed of hematopoietic marrow and marrow fat cells (*clear spaces*) attached to a thrombus. **B,** Amniotic fluid emboli. Two small pulmonary arterioles are packed with laminated swirls of fetal squamous cells. The surrounding lung is edematous and congested.
(Courtesy of Dr. Beth Schwartz, Baltimore, Maryland.)

The underlying cause is entry of amniotic fluid (and its contents) into the maternal circulation via tears in the placental membranes and/or uterine vein rupture. Histologic analysis reveals squamous cells shed from fetal skin, lanugo hair, fat from vernix caseosa, and mucin derived from the fetal respiratory or gastrointestinal tracts in the maternal pulmonary microcirculation (Fig. 13-16, *B*). Other findings include marked pulmonary edema, diffuse alveolar damage (Chapter 12), and systemic fibrin thrombi generated by disseminated intravascular coagulation.

Air Embolism

Gas bubbles within the circulation can coalesce and obstruct vascular flow and cause distal ischemic injury. Thus, a small volume of air trapped in a coronary artery during bypass surgery or introduced into the cerebral arterial circulation by neurosurgery performed in an upright "sitting position" can occlude flow, with dire consequences. Small

venous gas emboli generally have no deleterious effects, but sufficient air can enter the pulmonary circulation inadvertently during obstetric procedures or as a consequence of a chest wall injury to cause hypoxia, and very large venous emboli may arrest in the heart and cause death.

A particular form of gas embolism called *decompression sickness* is caused by sudden changes in atmospheric pressure. Thus, scuba divers, underwater construction workers, and persons in unpressurized aircraft who undergo rapid ascent are at risk. When air is breathed at high pressure (e.g., during a deep sea dive), increased amounts of gas (particularly nitrogen) become dissolved in the blood and tissues. If the diver then ascends (depressurizes) too rapidly, the nitrogen expands in the tissues and bubbles out of solution in the blood to form gas emboli, which cause tissue ischemia. Rapid formation of gas bubbles within skeletal muscles and supporting tissues in and about joints is responsible for the painful condition called "the bends" (so named in the 1880s because the afflicted person arches the back in a manner reminiscent of a then-popular women's fashion pose called the *Grecian bend*). Gas bubbles in the pulmonary vasculature cause edema, hemorrhages, and focal atelectasis or emphysema, leading to respiratory distress, the so-called *chokes.* A more chronic form of decompression sickness is called *caisson disease* (named for pressurized underwater vessels used during bridge construction) in which recurrent or persistent gas emboli in the bones lead to multifocal ischemic necrosis; the heads of the femurs, tibiae, and humeri are most commonly affected.

Acute decompression sickness is treated by placing affected persons in a high-pressure chamber, to force the gas back into solution. Subsequent slow decompression permits gradual gas resorption and exhalation so that obstructive bubbles do not re-form.

SUMMARY

Embolism

- An embolus is a solid, liquid, or gaseous mass carried by the blood to a site distant from its origin; most are dislodged thrombi.
- Pulmonary emboli derive primarily from lower-extremity deep vein thrombi; their effects depend mainly on the size of the embolus and the location in which it lodges. Consequences may include right-sided heart failure, pulmonary hemorrhage, pulmonary infarction, or sudden death.
- Systemic emboli derive primarily from cardiac mural or valvular thrombi, aortic aneurysms, or atherosclerotic plaques; whether an embolus causes tissue infarction depends on the site of embolization and the presence or absence of collateral circulation.

INFARCTION

An infarct is an area of ischemic necrosis caused by occlusion of the vascular supply to the affected tissue; the process by which such lesions form termed *infarction,* is a common and extremely important cause of clinical illness. Roughly 40% of all deaths in the United States are a consequence of cardiovascular disease, with most of these deaths stemming from myocardial or cerebral infarction. Pulmonary infarction is a common clinical complication, bowel infarction often is fatal, and ischemic necrosis of distal extremities (*gangrene*) causes substantial morbidity in the diabetic population.

Arterial thrombosis or arterial embolism underlies the vast majority of infarctions. Less common causes of arterial obstruction include vasospasm, expansion of an atheroma secondary to intraplaque hemorrhage, and extrinsic compression of a vessel, such as by tumor, a dissecting aortic aneurysm, or edema within a confined space (e.g., in *anterior tibial compartment syndrome*). Other uncommon causes of tissue infarction include vessel twisting (e.g., in testicular torsion or bowel volvulus), traumatic vascular rupture, and entrapment in a hernia sac. Although venous thrombosis can cause infarction, the more common outcome is simply congestion; typically, bypass channels rapidly open to provide sufficient outflow to restore the arterial inflow. Infarcts caused by venous thrombosis thus usually occur only in organs with a single efferent vein (e.g., testis or ovary).

MORPHOLOGY

Infarcts are classified on the basis of their color (reflecting the amount of hemorrhage) and the presence or absence of microbial infection. Thus, infarcts may be either **red (hemorrhagic)** or **white (anemic)** and may be either **septic** or **bland.**

Red infarcts (Fig. 3–17, *A*) occur (1) with venous occlusions (such as in ovarian torsion); (2) in loose tissues (e.g., lung) where blood can collect in infarcted zones; (3) in tissues with **dual circulations** such as lung and small intestine, where partial, inadequate perfusion by collateral arterial supplies is typical; (4) in previously congested tissues (as a consequence of sluggish venous outflow); and (5) when flow is reestablished after infarction has occurred (e.g., after angioplasty of an arterial obstruction).

White infarcts occur with arterial occlusions in solid organs with end-arterial circulations (e.g., heart, spleen, and kidney), and where tissue density limits the seepage of blood from adjoining patent vascular beds (Fig. 3–17, *B*). Infarcts tend to be wedge-shaped, with the occluded vessel at the apex and the organ periphery forming the base (Fig. 3–17); when the base is a serosal surface, there is often an overlying fibrinous exudate. Lateral margins may be irregular, reflecting flow from adjacent vessels. The margins of acute infarcts typically are indistinct and slightly hemorrhagic; with time, the edges become better defined by a narrow rim of hyperemia attributable to inflammation.

Infarcts resulting from arterial occlusions in organs without a dual circulation typically become progressively paler and sharply defined with time (Fig. 3–17, *B*). By comparison, hemorrhagic infarcts are the rule in the lung and other spongy organs (Fig. 3–17, *A*). Extravasated red cells in hemorrhagic infarcts are phagocytosed by macrophages, and the heme iron is converted to intracellular hemosiderin. Small amounts

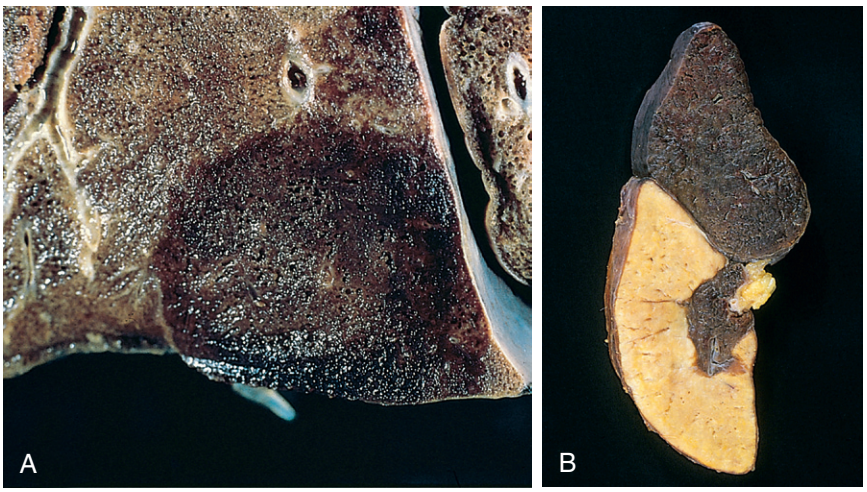


Figure 3–17 Red and white infarcts. **A,** Hemorrhagic, roughly wedge-shaped pulmonary infarct (red infarct). **B,** Sharply demarcated pale infarct in the spleen (white infarct).

do not impart any appreciable color to the tissue, but extensive hemorrhages leave a firm, brown residuum.

In most tissues, the main histologic finding associated with infarcts is **ischemic coagulative necrosis** (Chapter 1). An inflammatory response begins to develop along the margins of infarcts within a few hours and usually is well defined within 1 to 2 days. Eventually, inflammation is followed by repair, beginning in the preserved margins (Chapter 2). In some tissues, parenchymal regeneration can occur at the periphery of the infarct, where the underlying stromal architecture has been spared. Most infarcts, however, are ultimately replaced by scar (Fig. 3–18). The brain is an exception to these generalizations: Ischemic tissue injury in the central nervous system results in **liquefactive necrosis** (Chapter 1).

Septic infarctions occur when infected cardiac valve vegetations embolize, or when microbes seed necrotic tissue. In these cases the infarct is converted into an **abscess,** with a correspondingly greater inflammatory response (Chapter 2).

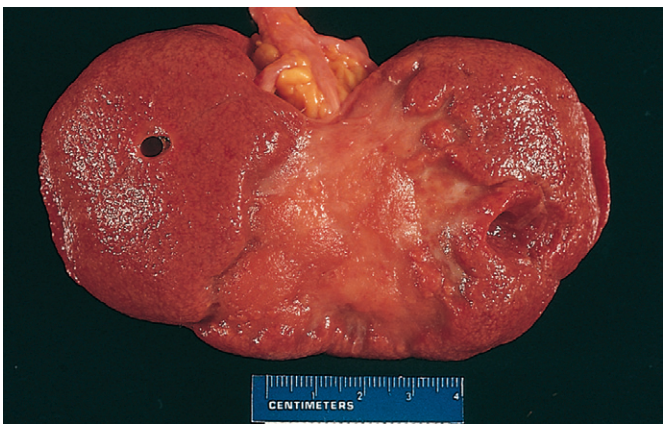


Figure 3–18 Remote kidney infarct, now replaced by a large fibrotic scar.

Factors That Influence Infarct Development. The effects of vascular occlusion range from inconsequential to tissue necrosis leading to organ dysfunction and sometimes death. The range of outcomes is influenced by (1) the anatomy of the vascular supply; (2) the time over which the occlusion develops; (3) the intrinsic vulnerability of the affected tissue to ischemic injury; and (4) the blood oxygen content.

- *Anatomy of the vascular supply.* The presence or absence of an alternative blood supply is the most important factor in determining whether occlusion of an individual vessel causes damage. The dual supply of the lung by the pulmonary and bronchial arteries means that obstruction of the pulmonary arterioles does not cause lung infarction unless the bronchial circulation also is compromised. Similarly, the liver, which receives blood from the hepatic artery and the portal vein, and the hand and forearm, with its parallel radial and ulnar arterial supply, are resistant to infarction. By contrast, the kidney and the spleen both have end-arterial circulations, and arterial obstruction generally leads to infarction in these tissues.
- *Rate of occlusion.* Slowly developing occlusions are less likely to cause infarction because they allow time for the development of collateral blood supplies. For example, small interarteriolar anastomoses, which normally carry minimal blood flow, interconnect the three major coronary arteries. If one coronary artery is slowly occluded (e.g., by encroaching atherosclerotic plaque), flow in this *collateral circulation* may increase sufficiently to prevent infarction—even if the original artery becomes completely occluded.
- *Tissue vulnerability to ischemia.* Neurons undergo irreversible damage when deprived of their blood supply for only 3 to 4 minutes. Myocardial cells, although hardier than neurons, still die after only 20 to 30 minutes of ischemia. By contrast, fibroblasts within myocardium remain viable after many hours of ischemia.
- *Hypoxemia.* Understandably, abnormally low blood O_2 content (regardless of cause) increases both the likelihood and extent of infarction.

SUMMARY

Infarction

- Infarcts are areas of ischemic necrosis most commonly caused by arterial occlusion (typically due to thrombosis or embolization); venous outflow obstruction is a less frequent cause.
- Infarcts caused by venous occlusion or occurring in spongy tissues typically are hemorrhagic (red); those caused by arterial occlusion in compact tissues typically are pale (white).
- Whether or not vascular occlusion causes tissue infarction is influenced by collateral blood supplies, the rate at which an obstruction develops, intrinsic tissue susceptibility to ischemic injury, and blood oxygenation.

SHOCK

Shock is the final common pathway for several potentially lethal events, including exsanguination, extensive trauma or burns, myocardial infarction, pulmonary embolism, and sepsis. Regardless of cause, *shock is characterized by systemic hypoperfusion of tissues; it can be caused by diminished cardiac output or by reduced effective circulating blood volume.* The consequences are *impaired tissue perfusion and cellular hypoxia.* Although shock initially is reversible, prolonged shock eventually leads to irreversible tissue injury that often proves fatal.

The most common forms of shock can be grouped into three pathogenic categories (Table 3–3):

- *Cardiogenic shock* results from low cardiac output due to myocardial pump failure. It may be caused by myocardial damage (infarction), ventricular arrhythmias, extrinsic compression (cardiac tamponade) (Chapter 10), or outflow obstruction (e.g., pulmonary embolism).
- *Hypovolemic shock* results from low cardiac output due to loss of blood or plasma volume (e.g., due to hemorrhage or fluid loss from severe burns).
- *Septic shock* results from arterial vasodilation and venous blood pooling that stems from the systemic immune response to microbial infection. Its complex pathogenesis is discussed in greater detail next.

Less commonly, shock can result from loss of vascular tone associated with anesthesia or secondary to a spinal cord injury (*neurogenic shock*). *Anaphylactic shock* results from systemic vasodilation and increased vascular permeability that is triggered by an immunoglobulin E–mediated hypersensitivity reaction (Chapter 4).

Pathogenesis of Septic Shock

Despite medical advances over the past several decades, septic shock remains a daunting clinical problem. Septic shock kills 20% of its victims, accounts for over 200,000 deaths annually in the United States, and is the number one cause of mortality in intensive care units. The incidence is rising, ironically, in part because of improved life support for critically ill patients, as well as an increase in invasive procedures and the growing numbers of immunocompromised patients (due to chemotherapy, immunosuppression, or HIV infection).

In septic shock, systemic arterial and venous dilation leads to tissue hypoperfusion, even though cardiac output is preserved or even initially increased. The decreased vascular tone is accompanied by widespread endothelial cell activation, often triggering a hypercoagulable state manifesting as disseminated intravascular coagulation. In addition, septic shock is associated with perturbations of metabolism that directly suppress cell and tissue function. *The net effect of these abnormalities is hypoperfusion and dysfunction of multiple organs.*

At present, gram-positive bacteria constitute the most common cause of septic shock, followed by gram-negative organisms and fungi. Although it was for a time thought that infections had to be disseminated to cause septic shock, infections localized to a specific tissue can trigger sepsis, even without detectable spread to the bloodstream. The ability of diverse flora to precipitate septic shock is consistent with the idea that several different microbial constituents can initiate the process. Most notably, macrophages, neutrophils, dendritic cells, endothelial cells, as well as soluble components of the innate immune system (e.g., complement) recognize and are activated by a variety of substances derived from microorganisms. Once activated, these cells and soluble factors initiate a number of inflammatory responses that interact in a complex, incompletely understood fashion to produce septic shock (Fig. 3–19). As an aside, a similar widespread inflammatory

Table 3–3 Three Major Types of Shock

Type of Shock	Clinical Examples	Principal Pathogenic Mechanisms
Cardiogenic	Myocardial infarction Ventricular rupture Arrhythmia Cardiac tamponade Pulmonary embolism	Failure of myocardial pump resulting from intrinsic myocardial damage, extrinsic pressure, or obstruction to outflow
Hypovolemic	Hemorrhage Fluid loss (e.g., vomiting, diarrhea, burns, trauma)	Inadequate blood or plasma volume
Septic	Overwhelming microbial infections Endotoxic shock Gram-positive septicemia Fungal sepsis Superantigens (e.g., toxic shock syndrome)	Peripheral vasodilation and pooling of blood; endothelial activation/injury; leukocyte-induced damage; disseminated intravascular coagulation; activation of cytokine cascades

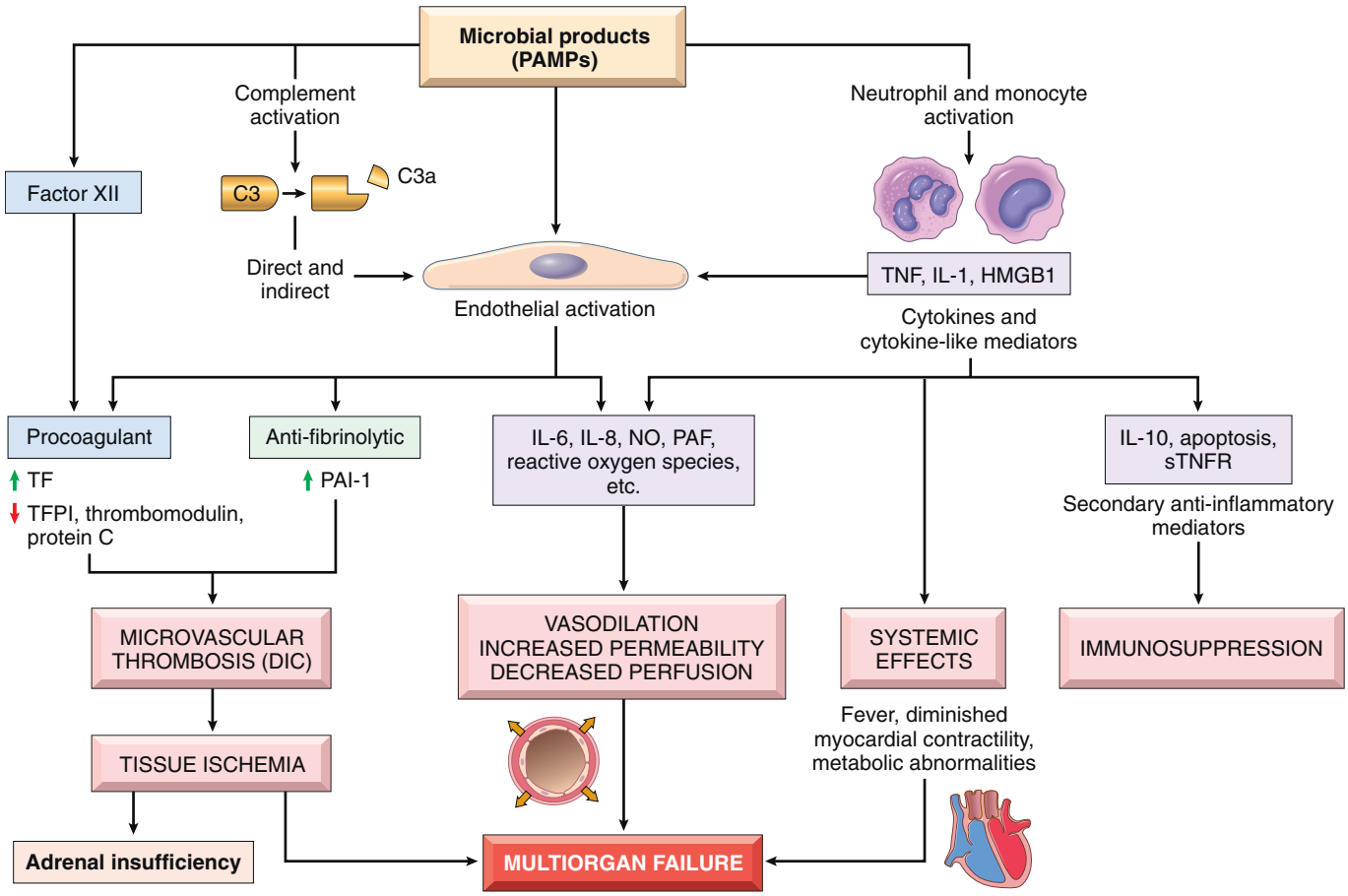


Figure 3–19 Major pathogenic pathways in septic shock. Microbial products activate endothelial cells and cellular and humoral elements of the innate immune system, initiating a cascade of events that lead to end-stage multiorgan failure. Additional details are given in the text. DIC, disseminated intravascular coagulation; HMGB1, high-mobility group box 1 protein; NO, nitric oxide; PAF, platelet-activating factor; PAI-1, plasminogen activator inhibitor-1; PAMP, pathogen-associated molecular pattern; STNFR, soluble tumor necrosis factor receptor; TF, tissue factor; TFPI, tissue factor pathway inhibitor.

response—the so-called *systemic inflammatory response syndrome (SIRS)*—can also be triggered in the absence of any apparent underlying infection; causes include extensive trauma or burns, pancreatitis, and diffuse ischemia.

Factors contributing to the pathophysiology of septic shock include the following:

- *Inflammatory mediators.* Cells of the innate immune system express receptors (e.g., Toll-like receptors [TLRs]) (Chapter 2) that recognize a host of microbe-derived substances containing so-called *pathogen-associated molecular patterns* (PAMPs). *Activation of pathogen recognition receptors by PAMPs triggers the innate immune responses that drive sepsis*. Upon activation, the inflammatory cells produce TNF and IL-1 (and other cytokines), plus cytokine-like mediators such as high-mobility group box 1 (HMGB1). Reactive oxygen species and lipid mediators such as prostaglandins and platelet-activating factor (PAF) also are elaborated (Chapter 2). These effector molecules activate endothelial cells, resulting in expression of adhesion molecules, a procoagulant phenotype, and secondary waves of cytokine production. The *complement cascade* also is activated by microbial components, both directly and through the proteolytic activity of plasmin (Chapter 2), resulting in the production of anaphylotoxins (C3a, C5a), chemotaxic fragments (C5a), and opsonins (C3b), all of which can contribute to the pro-inflammatory state.
- *Endothelial cell activation and injury.* Endothelial activation by microbial constituents or inflammatory cell mediators has three major sequelae: (1) thrombosis; (2) increased vascular permeability; and (3) vasodilation.

  The derangement in coagulation is sufficient to produce the formidable complication of disseminated intravascular coagulation in up to half of septic patients. Sepsis alters the expression of many factors ultimately favoring coagulation. Pro-inflammatory cytokines result in increased tissue factor production, while at the same time dampening fibrinolysis by increasing PAI expression. The production of other endothelial anticoagulant factors, such as tissue factor pathway inhibitor, thrombomodulin, and protein C, is also diminished. The procoagulant tendency is further enhanced by decreased blood flow within small vessels, which produces stasis and diminishes the washout of activated coagulation factors. Acting in concert, these effects promote the systemic deposition of fibrin-rich thrombi in small vessels, thus exacerbating tissue hypoperfusion. In full-blown disseminated intravascular coagulation, there

is also *consumption* of clotting factors and platelets, leading to concomitant bleeding and hemorrhage (Chapter 11).

The pro-inflammatory state associated with sepsis leads to widespread vascular leakage and tissue edema, with deleterious effects on both nutrient delivery and waste removal. It appears that inflammatory cytokines loosen endothelial cell tight junctions by causing the adhesion molecule VE-cadherin to be displaced from the junctions. The altered junctions become leaky, resulting in the accumulation of protein-rich exudates and edema throughout the body.

Expression of vasoactive inflammatory mediators (e.g., C3a, C5a, PAF), together with increased NO production, leads to systemic relaxation of vascular smooth muscle, producing hypotension and further reductions in tissue perfusion.

- *Metabolic abnormalities.* Septic patients exhibit insulin resistance and hyperglycemia. Cytokines such as TNF and IL-1, stress-induced hormones (such as glucagon, growth hormone, and glucocorticoid), and catecholamines all drive gluconeogenesis. At the same time, the pro-inflammatory cytokines suppress insulin release while simultaneously promoting insulin resistance in skeletal muscle and other tissues. Hyperglycemia suppresses neutrophil function—thereby decreasing bactericidal activity—and causes increased adhesion molecule expression on endothelial cells. Although sepsis initially is associated with a surge in glucocorticoid production, this increase is frequently followed by adrenal insufficiency and a relative glucocorticoid deficit. This effect may stem from depression of the synthetic capacity of adrenal glands or frank adrenal necrosis due to disseminated intravascular coagulation (*Waterhouse-Friderichsen syndrome*) (Chapter 19).
- *Immune suppression.* The hyperinflammatory state initiated by sepsis can paradoxically lead to a state of immunosuppression. Proposed mechanisms include production of anti-inflammatory mediators (e.g., soluble TNF receptor and IL-1 receptor antagonist), and widespread apoptosis of lymphocytes in the spleen and lymph nodes, the cause of which is uncertain. It is still debated whether immunosuppressive mediators are deleterious or protective in sepsis.
- *Organ dysfunction.* Systemic hypotension, increased vascular permeability, tissue edema, and small vessel thrombosis all decrease the delivery of oxygen and nutrients to the tissues and contribute to organ dysfunction. High levels of cytokines and secondary mediators can reduce myocardial contractility, thereby blunting cardiac output; increased vascular permeability and endothelial injury in the pulmonary circulation lead to the *acute respiratory distress syndrome* (ARDS) (Chapter 13). Ultimately, these factors conspire to cause multiorgan failure, particularly of the kidneys, liver, lungs, and heart, culminating in death.

Outcomes in patients with septic shock are difficult to predict; in general those with widespread infections and comorbid diseases have the highest mortality rates, but even young healthy individuals with virulent infections (e.g., meningococcal sepsis) can succumb within hours.

In view of the multiplicity of factors and the complexity of the interactions that underlie sepsis, it is perhaps not surprising that most attempts to intervene therapeutically with inhibitors of specific mediators have been of very modest benefit at best. The standard of care remains treatment with appropriate antibiotics, intensive insulin therapy for hyperglycemia, fluid resuscitation to maintain systemic pressures, and "physiologic doses" of corticosteroids to correct relative adrenal insufficiency. Some promising results have been observed in models of sepsis with treatments directed at restoring endothelial cell integrity.

An additional group of secreted bacterial proteins called *superantigens* also cause a syndrome similar to septic shock (e.g., *toxic shock syndrome*). Superantigens are polyclonal T-lymphocyte activators that induce T cells to release high levels of cytokines, which in turn results in a variety of clinical manifestations, ranging from a diffuse rash to vasodilation, hypotension, and death.

Stages of Shock

Shock is a progressive disorder that leads to death if the underlying problems are not corrected. The exact mechanisms of sepsis-related death are still unclear; aside from increased lymphocyte and enterocyte apoptosis, cellular necrosis is minimal. Death typically follows the failure of multiple organs, which usually offer no morphological clues to explain their dysfunction. For hypovolemic and cardiogenic shock, however, the pathways leading to a patient's demise are reasonably well understood. Unless the insult is massive and rapidly lethal (e.g., exsanguination from a ruptured aortic aneurysm), shock tends to evolve through three general (albeit somewhat artificial) stages. These stages have been documented most clearly in hypovolemic shock but are common to other forms as well:

- An initial *nonprogressive stage,* during which reflex compensatory mechanisms are activated and vital organ perfusion is maintained
- A *progressive stage,* characterized by tissue hypoperfusion and onset of worsening circulatory and metabolic derangement, including acidosis
- An *irreversible stage,* in which cellular and tissue injury is so severe that even if the hemodynamic defects are corrected, survival is not possible

In the early nonprogressive phase of shock, various *neurohumoral mechanisms* help maintain cardiac output and blood pressure. These mechanisms include baroreceptor reflexes, release of catecholamines and antidiuretic hormone, activation of the renin-angiotensin-aldersterone axis, and generalized sympathetic stimulation. The net effect is *tachycardia, peripheral vasoconstriction,* and *renal fluid conservation*; cutaneous vasoconstriction causes the characteristic "shocky" skin coolness and pallor (notably, septic shock can initially cause cutaneous *vasodilation,* so the patient may present with *warm, flushed skin*). Coronary and cerebral vessels are less sensitive to sympathetic signals and maintain relatively normal caliber, blood flow, and oxygen delivery. Thus, blood is shunted away from the skin to the vital organs such as the heart and the brain.

If the underlying causes are not corrected, shock passes imperceptibly to the progressive phase, which as noted is

characterized by widespread tissue hypoxia. In the setting of persistent oxygen deficit, intracellular aerobic respiration is replaced by anaerobic glycolysis with excessive production of lactic acid. The resultant metabolic *lactic acidosis lowers the tissue pH, which blunts the vasomotor response*; arterioles dilate, and blood begins to pool in the microcirculation. Peripheral pooling not only worsens the cardiac output but also puts endothelial cells at risk for the development of anoxic injury with subsequent DIC. With widespread tissue hypoxia, vital organs are affected and begin to fail.

In the absence of appropriate intervention, the process eventually enters an irreversible stage. Widespread cell injury is reflected in lysosomal enzyme leakage, further aggravating the shock state. Myocardial contractile function worsens, in part because of increased nitric oxide synthesis. The ischemic bowel may allow intestinal flora to enter the circulation, and thus bacteremic shock may be superimposed. Commonly, further progression to renal failure occurs as a consequence of ischemic injury of the kidney (Chapter 13), and despite the best therapeutic interventions, the downward spiral frequently culminates in death.

MORPHOLOGY

The cellular and tissue effects of shock are essentially those of hypoxic injury (Chapter 1) and are caused by a combination of **hypoperfusion and microvascular thrombosis.** Although any organ can be affected, brain, heart, kidneys, adrenals, and gastrointestinal tract are most commonly involved. **Fibrin thrombi** can form in any tissue but typically are most readily visualized in kidney glomeruli. **Adrenal cortical cell lipid depletion** is akin to that seen in all forms of stress and reflects increased utilization of stored lipids for steroid synthesis. While the lungs are resistant to hypoxic injury in hypovolemic shock occurring after hemorrhage, sepsis or trauma can precipitate diffuse alveolar damage (Chapter 12), leading to so-called **shock lung.** Except for neuronal and cardiomyocyte loss, affected tissues can recover completely if the patient survives.

Clinical Course

The clinical manifestations of shock depend on the precipitating insult. In hypovolemic and cardiogenic shock, patients exhibit hypotension, a weak rapid pulse, tachypnea, and cool, clammy, cyanotic skin. As already noted, in septic shock, the skin may be warm and flushed owing to peripheral vasodilation. The primary threat to life is the underlying initiating event (e.g., myocardial infarction, severe hemorrhage, bacterial infection). However, the cardiac, cerebral, and pulmonary changes rapidly aggravate the situation. If patients survive the initial period, worsening renal function can provoke a phase dominated by progressive oliguria, acidosis, and electrolyte imbalances.

Prognosis varies with the origin of shock and its duration. Thus, more than 90% of young, otherwise healthy patients with hypovolemic shock survive with appropriate management; by comparison, septic or cardiogenic shock is associated with substantially worse outcomes, even with state-of-the-art care.

SUMMARY

Shock

- Shock is defined as a state of systemic tissue hypoperfusion due to reduced cardiac output and/or reduced effective circulating blood volume.
- The major types of shock are cardiogenic (e.g., myocardial infarction), hypovolemic (e.g., blood loss), and septic (e.g., infections).
- Shock of any form can lead to hypoxic tissue injury if not corrected.
- Septic shock is caused by the host response to bacterial or fungal infections; it is characterized by endothelial cell activation, vasodilation, edema, disseminated intravascular coagulation, and metabolic derangements.

BIBLIOGRAPHY

Akhtar S: Fat embolism. Anesthesiol Clin 27:533, 2009. *[Recent overview of the pathogenesis and clinical issues in fat embolism syndrome.]*

Coppola A, Tufano A, Cerbone AM, Di Minno G: Inherited thrombophilia: implications for prevention and treatment of venous thromboembolism. Semin Thromb Hemost 35:683, 2009. *[Review of the genetic underpinnings of hypercoagulable states in a volume of the journal devoted to various aspects of thrombophilia.]*

Crawley J et al: The central role of thrombin in hemostasis. J Thromb Haemost 5 (Suppl 1):95, 2007. *[Review of the various pathways impacted by thrombin activation.]*

Crawley J, Lane D: The haemostatic role of tissue factor pathway inhibitor. Arterioscler Thromb Vasc Biol 28:233, 2008. *[Summary of the physiologic roles of TFPI.]*

Cushman M: Epidemiology and risk factors for venous thrombosis. Semin Hematol 44:62, 2007. *[Overview of the risk factors and pathophysiology of venous clotting.]*

Dahlback B: Blood coagulation and its regulation by anticoagulant pathways: genetic pathogenesis of bleeding and thrombotic diseases. J Intern Med 257:209, 2005. *[Although slightly older, this is a good one-stop review on normal and abnormal hemostasis.]*

Esmon CT, Esmon NL: The link between vascular features and thrombosis. Annu Rev Physiol 2011. *[Up-to-date review of the interactions of endothelium, blood flow, and hemostasis/thrombosis.]*

Goldhaber SZ: Advanced treatment strategies for acute pulmonary embolism, including thrombolysis and embolectomy. J Thromb Haemost 7(Suppl 1):322, 2009. *[Up-to-date guide to the recognition and therapy of pulmonary embolism.]*

Holy EW, Tanner FC: Tissue factor in cardiovascular disease pathophysiology and pharmacological intervention. Adv Pharmacol 59:259, 2010. *[Comprehensive review of the roles of tissue factor in hemostasis and potential pathways to be exploited in preventing pathologic thrombosis.]*

Hong MS, Amanullah AM: Heparin-induced thrombocytopenia: a practical review. Rev Cardiovasc Med 11:13, 2010. *[As characterized by the title, a good, practical review of the mechanisms and therapies for heparin-induced thrombocytopenia.]*

Hotchkiss R, Karl I: The pathophysiology and treatment of sepsis. N Engl J Med 348:138, 2003. *[Although an older paper, this is extremely well-written, and lays a solid pathogenic foundation for the pathways underlying sepsis.]*

Jennings LK: Mechanisms of platelet activation: need for new strategies to protect against platelet-mediated atherothrombosis. Thromb Haemost 102:248, 2009. *[Excellent and current review of the roles played by platelets in thrombosis and inflammation, as well as possible targets for therapeutic intervention.]*

Kwaan HC, Samama MM: The significance of endothelial heterogeneity in thrombosis and hemostasis. Semin Thromb Hemost 36:286, 2010. *[Review with newer data regarding the influence of endothelium on hemostasis and thrombosis.]*

Mackman N, Tilley RE, Key NS: Role of the extrinsic pathway of blood coagulation in hemostasis and thrombosis. Arterioscler Thromb Vasc Biol 27:1687, 2007. *[Good general overview of fundamental pathways in coagulation.]*

Montagnana M, Franchi M, Danese E, et al: Disseminated intravascular coagulation in obstetric and gynecologic disorders. Semin Thromb Hemost 36:404, 2010. *[Review of the mechanisms of DIC, and a good discussion of the pathophysiology of amniotic fluid embolism.]*

Munford RS: Severe sepsis and septic shock: the role of gram-negative bacteremia. Annu Rev Pathol 1:467, 2006. *[An interesting and provocative view regarding the pathogenesis of septic shock.]*

Osinbowale O, Ali L, Chi YW: Venous thromboembolism: a clinical review. Postgrad Med 122:54, 2010. *[Good review at a medical student/house officer level.]*

Stearns-Kurosawa DJ, Osuchowski MF, Valentine C, et al: The pathogenesis of sepsis. Ann Rev Pathol Mech Dis 6:19, 2011. *[Good update on some of the newer approaches to understanding and therapeutically targeting sepsis.]*

Rijken DC, Lijnen HR: New insights into the molecular mechanisms of the fibrinolytic system. J Thromb Haemost 7:4, 2009. *[Excellent review of fibrinolytic pathways.]*

Ruiz-Irastorza G, Crowther M, Branch W, Khamashta MA: Antiphospholipid syndrome. Lancet 376:1498, 2010. *[Good summary of the anti-phospholipid syndrome with emphasis on diagnosis and therapeutics.]*

Wu KK, Matijevic-Aleksic N: Molecular aspects of thrombosis and antithrombotic drugs. Crit Rev Clin Lab Sci 42:249, 2005. *[Lengthy, thorough overview of the mechanisms of thrombus formation with emphasis on targets for therapeutic intervention.]*

Zwicker J, Furie BC, Furie B: Cancer-associated thrombosis. Crit Rev Oncol Hematol 62:126, 2007. *[Thorough review of the mechanisms underlying the hypercoagulable state of malignancy.]*

See Targeted Therapy available online at studentconsult.com

CHAPTER 4

Diseases of the Immune System

CHAPTER CONTENTS

Immunity refers to protection against infections, and the immune system is the collection of cells and molecules that are responsible for defending the body against the countless pathogenic microbes in the environment. Deficiencies in immune defenses result in an increased susceptibility to infections, which can be life-threatening if the deficits are not corrected. On the other hand, the immune system is itself capable of causing great harm and is the root cause of some of the most vexing and intractable diseases of the modern world. Thus, diseases of immunity range from those caused by "too little" to those caused by "too much or inappropriate" immune activity.

This chapter starts with a brief review of some of the basic concepts of lymphocyte biology and normal immune responses, which establishes a foundation for the subsequent discussions of diseases caused by excessive or inappropriate immune responses, rejection of organ transplants and immune deficiency disorders. The chapter concludes with a discussion of amyloidosis, a disease characterized by the abnormal extracellular deposition of certain proteins (some of which are produced in the setting of immune responses).

INNATE AND ADAPTIVE IMMUNITY

Defense against microbes consists of two types of reactions (Fig. 4–1). *Innate immunity* (also called natural, or native, immunity) is mediated by cells and proteins that are always present and poised to fight against microbes, being called into action immediately in response to infection. The major components of innate immunity are epithelial barriers of the skin, gastrointestinal tract, and respiratory tract, which prevent microbe entry; phagocytic leukocytes (neutrophils and macrophages); a specialized cell type called the natural killer (NK) cell; and several circulating plasma proteins, the most important of which are the proteins of the complement system.

The innate immune response is able to prevent and control many infections. However, many pathogenic microbes have evolved to overcome the early defenses, and protection against these infections requires the more specialized and powerful mechanisms of *adaptive immunity* (also called acquired, or specific, immunity). Adaptive immunity is normally silent and responds (or "adapts") to

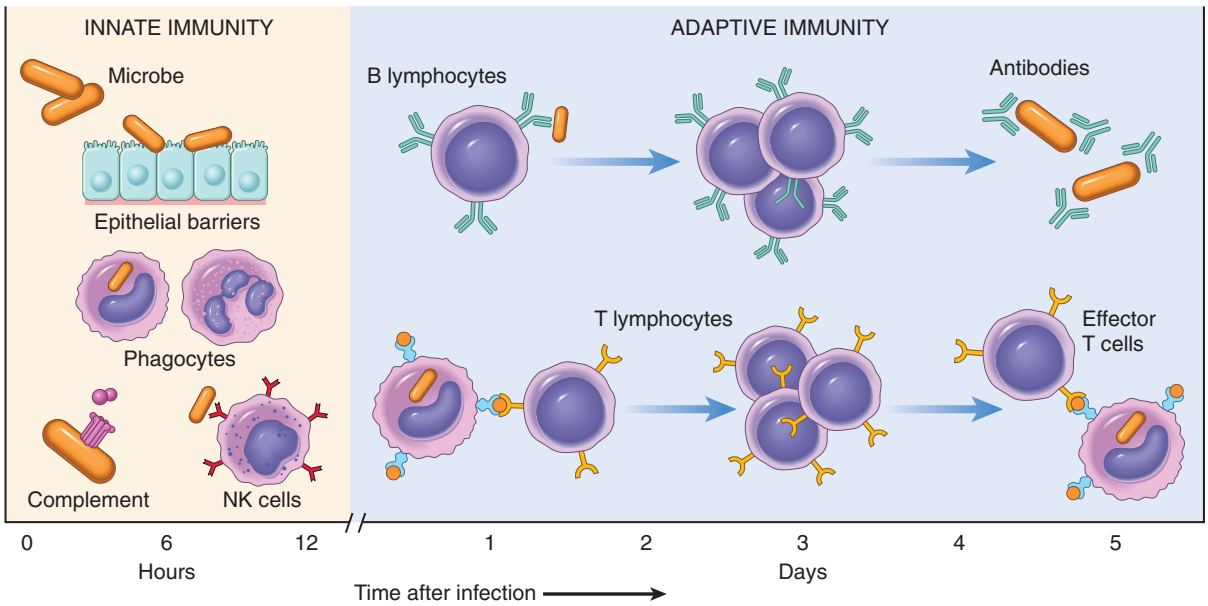


Figure 4–1 The principal mechanisms of innate immunity and adaptive immunity. NK, natural killer.

the presence of infectious microbes by becoming active, expanding, and generating potent mechanisms for neutralizing and eliminating the microbes. *The components of the adaptive immune system are lymphocytes and their products.* By convention, the terms "immune system" and "immune response" refer to adaptive immunity.

There are two types of adaptive immune responses: humoral immunity, mediated by soluble proteins called antibodies that are produced by B lymphocytes (also called B cells), and cell-mediated (or cellular) immunity, mediated by T lymphocytes (also called T cells). Antibodies provide protection against extracellular microbes in the blood, mucosal secretions, and tissues. T lymphocytes are important in defense against intracellular microbes. They work by either directly killing infected cells (accomplished by cytotoxic T lymphocytes) or by activating phagocytes to kill ingested microbes, via the production of soluble protein mediators called cytokines (made by helper T cells). The main properties and functions of the cells of the immune system are described in the next section.

When the immune system is inappropriately triggered or not properly controlled, the same mechanisms that are involved in host defense cause tissue injury and disease. The reaction of the cells of innate and adaptive immunity may be manifested as *inflammation.* As discussed in Chapter 2, inflammation is a beneficial process, but it is also the basis for many human diseases. Presented later in this chapter is an overview of the ways in which the adaptive immune response triggers pathologic inflammatory reactions.

CELLS AND TISSUES OF THE IMMUNE SYSTEM

The cells of the immune system consist of lymphocytes, which recognize antigens and mount adaptive immune responses; specialized antigen-presenting cells (APCs), which capture and display microbial and other antigens to the lymphocytes; and various effector cells, whose function is to eliminate microbes and other antigens. Two remarkable features of the immune system are the specialization of the cells to perform diverse functions, and the precise control mechanisms that permit useful responses when needed and prevent potentially harmful ones.

Lymphocytes

Lymphocytes are present in the circulation and in various lymphoid organs. Although all lymphocytes appear morphologically identical, there are actually several functionally and phenotypically distinct lymphocyte populations. Lymphocytes develop from precursors in the generative lymphoid organs; T lymphocytes are so called because they mature in the thymus, whereas B lymphocytes mature in the bone marrow. Each T or B lymphocyte expresses receptors for a single antigen, and the total population of lymphocytes (numbering about 10^{12} in humans) is capable of recognizing tens or hundreds of millions of antigens. This enormous diversity of antigen recognition is generated by the somatic rearrangement of antigen receptor genes during lymphocyte maturation, and variations that are introduced during the joining of different gene segments to form antigen receptors. These antigen receptors are rearranged and expressed in lymphocytes but not in any other cell. Therefore, the demonstration of antigen receptor gene rearrangements by molecular methods (e.g., polymerase chain reaction [PCR] assay) is a definitive marker of T or B lymphocytes. Because each lymphocyte has a unique DNA rearrangement (and hence a unique antigen receptor), molecular analysis of the rearrangements in cell populations can distinguish polyclonal (non-neoplastic) lymphocyte proliferations from monoclonal (neoplastic) expansions. Such analyses are used in the diagnosis of lymphoid malignancies (Chapter 11).

T Lymphocytes

Thymus-derived, or T, lymphocytes are the effector cells of cellular immunity and the "helper cells" for antibody responses to protein antigens. T cells constitute 60% to 70% of the lymphocytes in peripheral blood and are the major lymphocyte population in splenic periarteriolar sheaths and lymph node interfollicular zones. T cells do not detect free or circulating antigens. Instead, the vast majority (greater than 95%) of T cells recognize only peptide fragments of protein antigens bound to proteins of the major histocompatibility complex (MHC). The MHC was discovered on the basis of studies of graft rejection and acceptance (tissue, or "histo," compatibility). It is now known that *the normal function of MHC molecules is to display peptides for recognition by T lymphocytes.* By forcing T cells to see MHC-bound peptides on cell surfaces the system ensures that T cells can recognize antigens displayed by other cells. T cells function by interacting with other cells—either to kill infected cells or to activate phagocytes or B lymphocytes that have ingested protein antigens. In each person, T cells recognize only peptides displayed by that person's MHC molecules, which, of course, are the only MHC molecules that the T cells normally encounter. This phenomenon is called MHC restriction. Peptide antigens presented by self MHC molecules are recognized by the T cell receptor (TCR), which is a heterodimer composed of disulfide-linked α and β protein chains (Fig. 4–2, *A*); each chain has a variable region that participates in binding a particular peptide antigen and a constant region that interacts with associated signaling molecules.

TCRs are noncovalently linked to a cluster of five invariant polypeptide chains, the γ, δ, and ε proteins of the CD3 molecular complex and two ζ chains (Fig. 4–2, *A*). The CD3 proteins and ζ chains do not themselves bind antigens; instead, they are attached to the TCR and deliver intracellular biochemical signals after TCR recognition of antigen. In addition to these signaling proteins, T cells express a number of other invariant molecules that serve diverse functions. CD4 and CD8 are expressed on distinct T cell subsets and serve as coreceptors for T cell activation. During antigen recognition, CD4 molecules on T cells bind to invariant portions of class II MHC molecules (see later) on selected APCs; in an analogous fashion, CD8 binds to class I MHC molecules. CD4 is expressed on 50%–60% of mature T cells, whereas CD8 is expressed on about 40% of T cells. The CD4- and CD8-expressing T cells—called CD4+ and CD8+ cells, respectively—perform different but overlapping functions. CD4+ T cells are "helper" T cells because they secrete soluble molecules (cytokines) that help B cells to produce antibodies (the origin of the name "helper" cells) and also help macrophages to destroy phagocytosed microbes. The central role of CD4+ helper cells in immunity is highlighted by the severe compromise that results from the destruction of this subset by human immunodeficiency

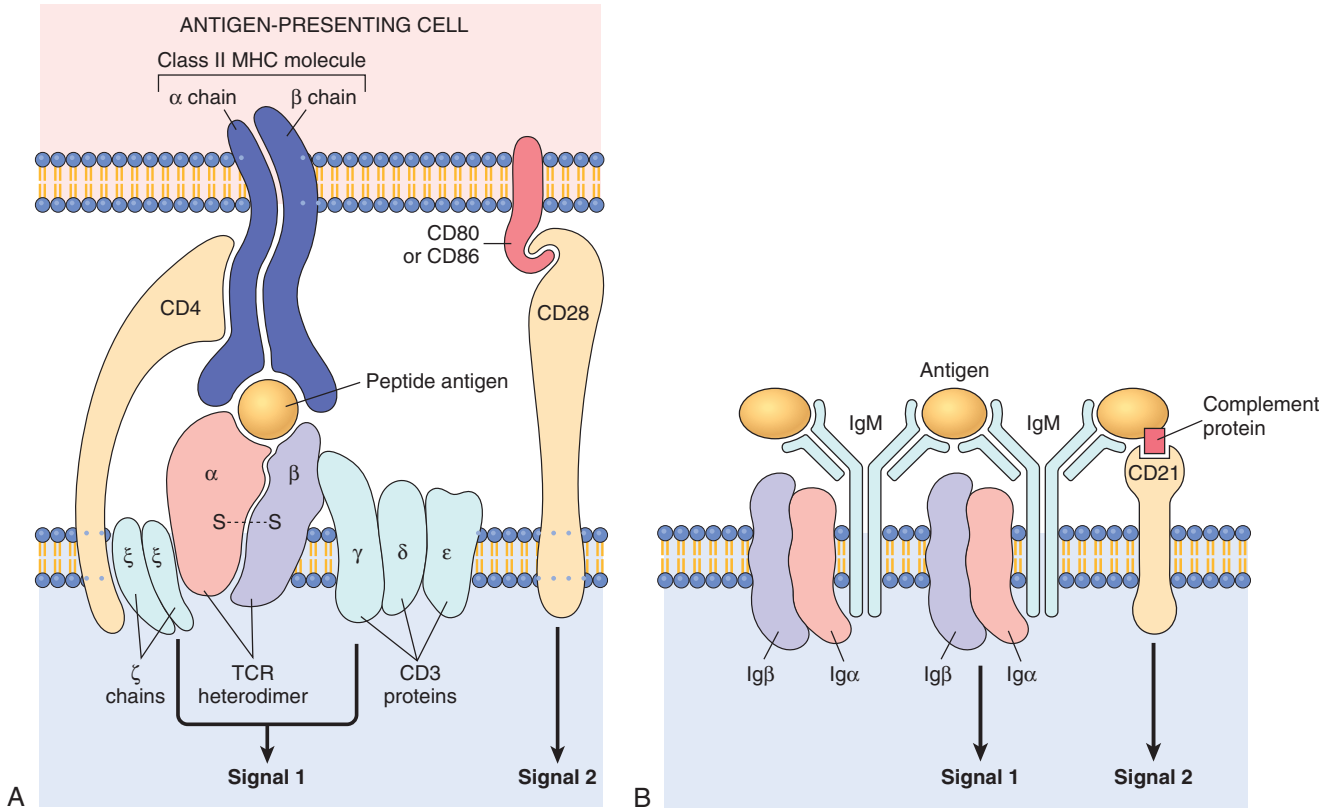


Figure 4–2 Lymphocyte antigen receptors. **A,** The T cell receptor (TCR) complex and other molecules involved in T cell activation. The TCRα and TCRβ chains recognize antigen (in the form of peptide–MHC complexes expressed on antigen-presenting cells), and the linked CD3 complex initiates activating signals. CD4 and CD28 are also involved in T cell activation. (Note that some T cells express CD8 and not CD4; these molecules serve analogous roles.) **B,** The B cell receptor complex is composed of membrane IgM (or IgD, *not shown*) and the associated signaling proteins Igα and Igβ. CD21 is a receptor for a complement component that promotes B cell activation. Ig, immunoglobulin; MHC, major histocompatibilty complex.

virus (HIV) infection. CD8+ T cells can also secrete cytokines, but they play a more important role in directly killing virus-infected or tumor cells, and hence are called "cytotoxic" T lymphocytes (CTLs). Other important invariant proteins on T cells include CD28, which functions as the receptor for molecules that are induced on APCs by microbes (and are called costimulators), and various adhesion molecules that strengthen the bond between the T cells and APCs and control the migration of the T cells to different tissues.

In a minority of peripheral blood T cells and in many of the T cells associated with mucosal surfaces (e.g., lung, gastrointestinal tract), the TCRs are heterodimers of γ and δ chains, which are similar but not identical to the α and β chains of most TCRs. Such γδ T cells, which do not express CD4 or CD8, recognize nonprotein molecules (e.g., bacterial lipoglycans), but their functional roles are not well understood. Another small population of T cells expresses markers of T cells and NK cells. These so-called NKT cells recognize microbial glycolipids, and may play a role in defense against some infections. The antigen receptors of NKT cells are much less diverse than the receptors of "conventional" T cells, suggesting that the former recognize conserved microbial structures.

Another population of T cells that functions to suppress immune responses is that of regulatory T lymphocytes. This cell type is described later, in the context of tolerance of self antigens.

Major Histocompatibility Complex Molecules: The Peptide Display System of Adaptive Immunity

Because MHC molecules are fundamental to T cell recognition of antigens, and because genetic variations in MHC molecules are associated with immunologic diseases, it is important to review the structure and function of these molecules. The human MHC, known as the human leukocyte antigen (HLA) complex, consists of a cluster of genes on chromosome 6 (Fig. 4–3). The HLA system is highly polymorphic; that is, there are several alternative forms (alleles) of a gene at each locus (estimated to number about 3500 for all HLA genes and about 1100 for HLA-B alleles alone). Such diversity provides a system whereby a vast range of peptides can be displayed by MHC molecules for recognition by T cells. As we shall see, this polymorphism also constitutes a formidable barrier to organ transplantation.

On the basis of their chemical structure, tissue distribution, and function, MHC gene products fall into two main categories:

- *Class I MHC* molecules are encoded by three closely linked loci, designated HLA-A, HLA-B, and HLA-C (Fig. 4–3). Each of these molecules is a heterodimer, consisting of a polymorphic 44-kDa α chain noncovalently associated with an invariant 12-kDa β_2-microglobulin polypeptide (encoded by a separate gene on chromosome 15). The extracellular portion of the α chain

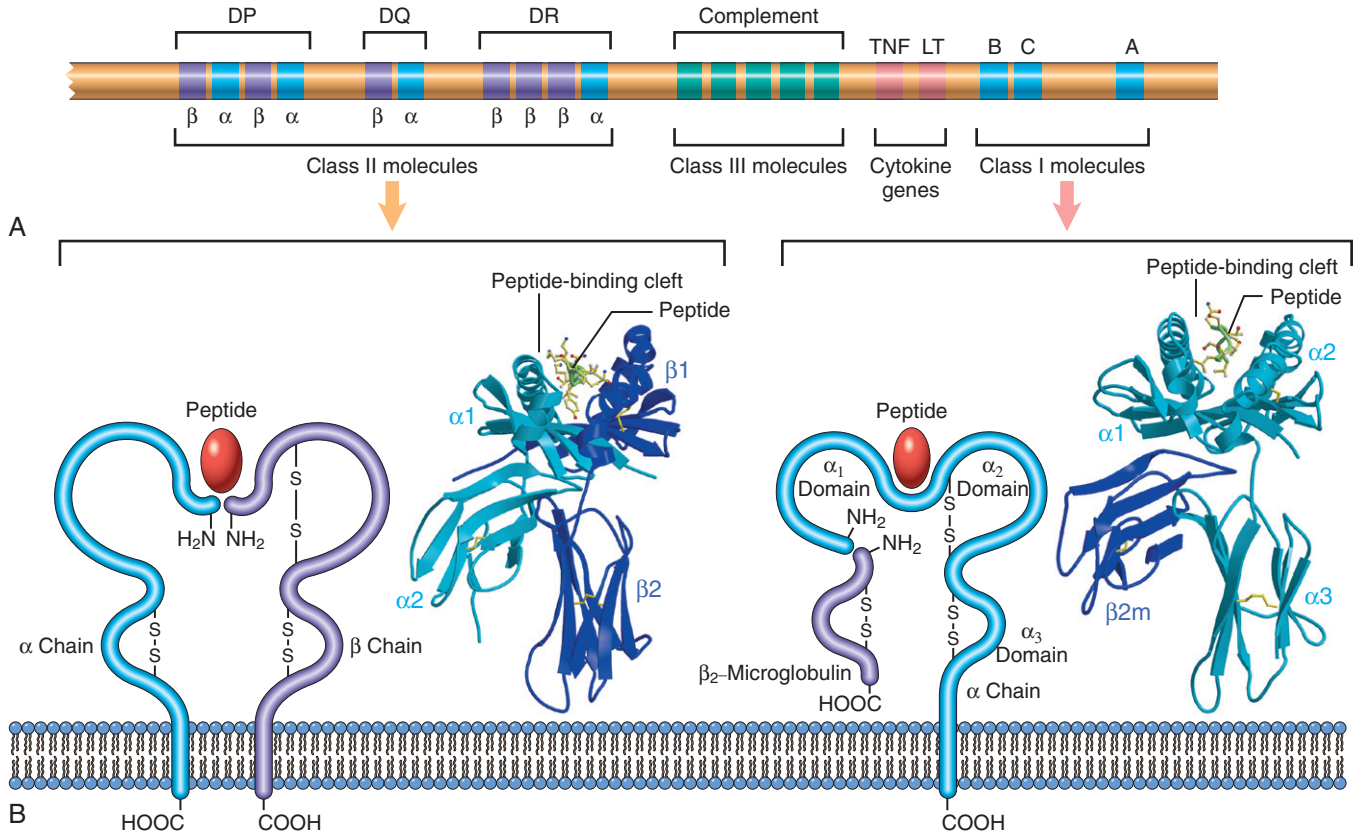


Figure 4–3 The human leukocyte antigen (HLA) complex and the structure of HLA molecules. **A,** The location of genes in the HLA complex. The sizes and distances between genes are not to scale. The class II region also contains genes that encode several proteins involved in antigen processing (not shown). **B,** Schematic diagrams and crystal structures of class I and class II HLA molecules. LT, lymphotoxin; TNF, tumor necrosis factor.
(Crystal structures are courtesy of Dr. P. Bjorkman, California Institute of Technology, Pasadena, California.)

contains a cleft where the polymorphic residues are located and where foreign peptides bind to MHC molecules for presentation to T cells, and a conserved region that binds CD8, ensuring that only CD8+ T cells can respond to peptides displayed by class I molecules. In general, class I MHC molecules bind and display peptides derived from proteins synthesized in the cytoplasm of the cell (e.g., viral antigens). Because class I MHC molecules are present on all nucleated cells, all virus-infected cells can be detected and eliminated by CD8+ CTLs.

- *Class II MHC* molecules are encoded by genes in the HLA-D region, which contains at least three subregions: DP, DQ, and DR. Class II MHC molecules are heterodimers of noncovalently linked polymorphic α and β subunits (Fig. 4–3). The extracellular portion of the class II MHC heterodimer contains a cleft for the binding of antigenic peptides and a region that binds CD4. Class II MHC expression is restricted to a few types of cells, mainly APCs (notably, dendritic cells [DCs]), macrophages, and B cells. In general, class II MHC molecules bind to peptides derived from proteins synthesized outside the cell (e.g., those derived from extracellular bacteria) and ingested into the cell. This property allows CD4+ T cells to recognize the presence of extracellular pathogens and to orchestrate a protective response.
- Several other proteins are encoded in the MHC locus, some of which have been called "class III molecules." These include complement components (C2, C3, and Bf) and the cytokines tumor necrosis factor (TNF) and lymphotoxin. These molecules do not form a part of the peptide display system and are not discussed further.

Each person inherits one HLA allele from each parent; typically, then, two different molecules are expressed for every HLA locus. Cells of a heterozygous person can therefore express six different class I HLA molecules: three of maternal origin and three of paternal origin. Similarly, a given individual expresses maternal and paternal alleles of the class II MHC loci; because some HLA-D α and β chains can mix and match with each other, each class II–expressing cell can have as many as 20 different class II MHC molecules. Different MHC alleles bind to different peptide fragments; *the expression of many different MHC molecules allows each cell to present a wide array of peptide antigens*.

As a result of the polymorphism at the major HLA loci in the population, a virtually infinite number of combinations of molecules exist, and each person expresses a unique MHC antigenic profile on his or her cells. The combination of HLA alleles for each person is called the HLA haplotype. The implications of HLA polymorphism are obvious in the context of transplantation—because each person has HLA alleles that differ to some extent from every other person's, grafts from virtually any donor will evoke immune responses in the recipient and be rejected (except, of course, for identical twins). In fact, HLA molecules were discovered in the course of early attempts at tissue transplantation. HLA molecules of the graft evoke both humoral and cell-mediated responses, eventually leading to graft destruction (as discussed later in this chapter). This ability of MHC molecules to trigger immune responses is the reason these molecules are often called "antigens." It is believed that the polymorphism of MHC genes arose to enable display of and response to any conceivable microbial peptide encountered in the environment.

The role of the MHC in T cell stimulation also has important implications for the genetic control of immune responses. The ability of any given MHC allele to bind the peptide antigens generated from a particular pathogen will determine whether a specific person's T cells can actually "see" and respond to that pathogen. The inheritance of particular alleles influences both protective and harmful immune responses. For example, if the antigen is ragweed pollen and the response is an allergic reaction, inheritance of some HLA genes may make individuals susceptible to "hay fever," the colloquial name for ragweed allergy. On the other hand, responsiveness to a viral antigen, determined by inheritance of certain HLA alleles, may be beneficial for the host.

Finally, *many autoimmune diseases are associated with particular HLA alleles*. We return to a discussion of HLA associations with diseases when we consider autoimmunity.

B Lymphocytes

Bone marrow–derived B lymphocytes are the cells that produce antibodies and are thus the effector cells of humoral immunity. B cells make up 10% to 20% of the circulating peripheral lymphocyte population. They also are present in bone marrow and in the follicles of peripheral lymphoid tissues (lymph nodes, spleen, tonsils, and other mucosal tissues).

B cells recognize antigen by means of membrane-bound antibody of the immunoglobulin M (IgM) class, expressed on the surface together with signaling molecules to form the B cell receptor (BCR) complex (Fig. 4–2, *B*). Whereas T cells can recognize only MHC-associated peptides, B cells can recognize and respond to many more chemical structures, including soluble or cell-associated proteins, lipids, polysaccharides, nucleic acids, and small chemicals; furthermore, B cells (and antibodies) recognize native (properly folded) forms of these antigens. As with TCRs, each antibody has a unique antigen specificity. The diversity of antibodies is generated during somatic rearrangements of immunoglobulin genes. B cells express several invariant molecules that are responsible for signal transduction and for activation of the cells (Fig. 4–2, *B*). Some are the signaling molecules attached to the BCR; another example is CD21 (also known as the type 2 complement receptor, or CR2), which recognizes a complement breakdown product that frequently is deposited on microbes and promotes B cell responses to microbial antigens. Interestingly, the ubiquitous Epstein-Barr virus has cleverly evolved to use CD21 as a receptor for binding to B cells and infecting them.

After stimulation, B cells differentiate into *plasma cells,* which secrete large amounts of antibodies, the mediators of humoral immunity. There are five classes, or isotypes, of immunoglobulins: IgG, IgM, and IgA constitute more than 95% of circulating antibodies. IgA is the major isotype in mucosal secretions; IgE is present in the circulation at very low concentrations and also is found attached to the surfaces of tissue mast cells; and IgD is expressed on the surfaces of B cells but is not secreted. As discussed later, each isotype has characteristic abilities to activate

complement or recruit inflammatory cells and thus plays a different role in host defense and disease states.

Natural Killer Cells

Natural killer (NK) cells are lymphocytes that arise from the common lymphoid progenitor that gives rise to T and B lymphocytes. However, NK cells are cells of innate immunity and do not express highly variable and clonally distributed receptors for antigens. Therefore, *they do not have specificities as diverse as do T cells or B cells*. NK cells have two types of receptors—inhibitory and activating. The inhibitory receptors recognize self class I MHC molecules, which are expressed on all healthy cells, whereas the activating receptors recognize molecules that are expressed or upregulated on stressed or infected cells or cells with DNA damage. Normally, the effects of the inhibitory receptors dominate over those of the activating receptors, thereby preventing activation of the NK cells. Infections (especially viral infections) and stress are associated with reduced expression of class I MHC molecules, thus releasing the NK cells from inhibition. At the same time, there is increased engagement of the activating receptors. The net result is that the NK cells are activated and the infected or stressed cells are killed and eliminated.

Antigen-Presenting Cells

The immune system contains several cell types that are specialized to capture microbial antigens and display these to lymphocytes. Foremost among these APCs are dendritic cells (DCs), the major cells for displaying protein antigens to naive T cells to initiate immune responses. Several other cell types present antigens to different lymphocytes at various stages of immune responses.

Dendritic Cells

Cells with dendritic morphology (i.e., with fine dendritic cytoplasmic processes) occur as two functionally distinct types. *Dendritic cells (DCs)*, sometimes called interdigitating DCs, express high levels of class II MHC and T cell costimulatory molecules and function to capture and present antigens to T cells. DCs reside in and under epithelia, where they are strategically located to capture entering microbes; an example is the Langerhans cell of the epidermis. DCs also are present in the T cell zones of lymphoid tissues, where they present antigens to T cells circulating through these tissues, and in the interstitium of many nonlymphoid organs, such as the heart and lungs, where they are poised to capture the antigens of any invading microbes. One subset of DCs is called *plasmacytoid DCs* because of their resemblance to plasma cells. These cells are present in the blood and lymphoid organs, and are major sources of the antiviral cytokine type I interferon, produced in response to many viruses.

The second type of cells with dendritic morphology are *follicular dendritic cells (FDCs)*. These cells are located in the germinal centers of lymphoid follicles in the spleen and lymph nodes. FDCs bear receptors for the Fc tails of IgG molecules and for complement proteins and hence efficiently trap antigens bound to antibodies and complement. These cells display antigens to activated B lymphocytes in lymphoid follicles and promote secondary antibody responses, but are not involved in capturing antigens for display to T cells.

Other Antigen-Presenting Cells

Macrophages ingest microbes and other particulate antigens and display peptides for recognition by T lymphocytes. These T cells in turn activate the macrophages to kill the microbes, the central reaction of cell-mediated immunity. B cells present peptides to helper T cells and receive signals that stimulate antibody responses to protein antigens.

Effector Cells

Many different types of leukocytes perform the ultimate task of the immune response, which is to eliminate infections. NK cells are front-line effector cells in that they can rapidly react against "stressed" cells. Antibody-secreting plasma cells are the effector cells of humoral immunity. T lymphocytes, both CD4+ helper T cells and CD8+ CTLs, are effector cells of cell-mediated immunity. These lymphocytes often function in host defense together with other cells. Macrophages, as described in Chapter 2, bind microbes that are coated with antibodies or complement and then phagocytose and destroy these microbes, thus serving as effector cells of humoral immunity. Macrophages also respond to signals from helper T cells, which improves their ability to destroy phagocytosed microbes, thus serving as effector cells of cellular immunity. T lymphocytes secrete cytokines that recruit and activate other leukocytes, such as neutrophils and eosinophils, and together these cell types function in defense against various pathogens.

Lymphoid Tissues

The lymphoid tissues of the body are divided into generative (primary) organs, where lymphocytes express antigen receptors and mature, and peripheral (secondary) lymphoid organs, where adaptive immune responses develop. The generative organs are the thymus and bone marrow, and the peripheral organs are the lymph nodes, spleen, and mucosal and cutaneous lymphoid tissues. Mature lymphocytes recirculate through the peripheral organs, hunting for microbial antigens that they can respond to. An important characteristic of these organs is that T and B lymphocytes are anatomically organized in a manner that facilitates the adaptive immune response, a process that is described later.

SUMMARY

Cells and Tissues of the Immune System

- Lymphocytes are the mediators of adaptive immunity and the only cells that produce specific and diverse receptors for antigens.
- T (thymus-derived) lymphocytes express TCRs that recognize peptide antigens displayed by MHC molecules on the surface of APCs.

- B (bone marrow–derived) lymphocytes express membrane-bound antibodies that recognize a wide variety of antigens. B cells are activated to become plasma cells, which secrete antibodies.
- NK cells kill cells that are infected by some microbes or are stressed and damaged beyond repair. NK cells express inhibitory receptors that recognize MHC molecules that are normally expressed on healthy cells, and are thus prevented from killing normal cells.
- APCs capture microbes and other antigens, transport them to lymphoid organs, and display them for recognition by lymphocytes. The most efficient APCs are DCs, which are located in epithelia and most tissues.
- The cells of the immune system are organized in tissues. Some of these tissues are the sites of mature lymphocyte production (the generative lymphoid organs, the bone marrow and thymus), while others are the sites of immune responses (the peripheral lymphoid organs, including lymph nodes, spleen, and mucosal lymphoid tissues).

OVERVIEW OF NORMAL IMMUNE RESPONSES

The previous section described the major components of the immune system. This section summarizes the key features of normal immune responses. This overview will serve as a foundation for the subsequent discussions of diseases caused by deficient or uncontrolled immune responses.

The Early Innate Immune Response to Microbes

The principal barriers between hosts and their environment are the epithelia of the skin and the gastrointestinal and respiratory tracts. Infectious microbes usually enter through these routes and attempt to colonize the hosts. The mechanisms of innate immunity operate at every step in a microbe's attempt to invade. At the site of entry, epithelia serve as physical barriers to infections and eliminate microbes through production of peptide antibiotics and the actions of intraepithelial lymphocytes. If microbes are able to survive and traverse these epithelia, they encounter phagocytes, including neutrophils, which are rapidly recruited from the blood into tissues, and macrophages, which live in tissues under epithelia. The function of these phagocytic cells is to ingest microbes and destroy them by producing microbicidal substances. In response to recognition of microbes, phagocytes, DCs, and many other cell types secrete proteins called cytokines (described later), which promote inflammation and microbial killing and enhance protective immune responses. Cells use several receptors to sense microbes; foremost among these are the Toll-like receptors (TLRs), so named because of homology with the *Drosophila* Toll protein, that recognize bacterial and viral components (Chapter 2). NK cells kill virus-infected cells and produce the macrophage-activating cytokine IFN-γ. If the microbes enter the blood, many plasma proteins, including the proteins of the complement system, recognize the microbes and are activated, and their products kill microbes and coat (opsonize) the microbes for phagocytosis. In addition to combating infections, innate immune responses stimulate subsequent adaptive immunity, providing signals that are essential for initiating the responses of antigen-specific T and B lymphocytes.

The Capture and Display of Microbial Antigens

Microbes that enter through epithelia, along with their protein antigens, are captured by DCs that are resident in and under these epithelia. Antigen-bearing DCs then migrate to draining lymph nodes (Fig. 4–4). Protein antigens are proteolytically digested in the APCs to generate peptides that are displayed on the surface of the APCs bound to MHC molecules. Antigens in different cellular compartments are presented by different MHC molecules and are recognized by different subsets of T cells. Antigens that are ingested from the extracellular environment are processed in endosomal and lysosomal vesicles and then are displayed bound to class II MHC molecules. Because CD4 binds to class II MHC molecules, CD4+ helper T cells recognize class II–associated peptides. By contrast, antigens in the cytoplasm are displayed by class I MHC molecules and are recognized by CD8+ cytotoxic T cells, because CD8 binds to class I MHC. This segregation of different antigens is key to the specialized functions of CD4+ and CD8+ T cells; as we discuss below, the two classes of T cells are designed to combat microbes that are located in different cellular compartments. Protein antigens, as well as polysaccharides and other nonprotein antigens, can also be recognized directly by B lymphocytes in the lymphoid follicles of the peripheral lymphoid organs.

Before being recognized by B and T cells, the microbe elicits an innate immune response. This response activates APCs to express costimulatory molecules and secrete cytokines that stimulate the proliferation and differentiation of T lymphocytes. The principal costimulators for T cells are the B7 molecules (CD80 and CD86) that are expressed on APCs and recognized by the CD28 receptor on naive T cells. The innate immune response to some microbes and polysaccharides also results in the activation of complement, generating cleavage products that enhance the proliferation and differentiation of B lymphocytes. Thus, antigen (signal 1 in Fig. 4–2) and molecules produced during innate immune responses (signal 2 in Fig. 4–2) function cooperatively to activate antigen-specific lymphocytes. The requirement for microbe-triggered signal 2 ensures that the adaptive immune response is induced by microbes and not by harmless substances.

Cell-Mediated Immunity: Activation of T Lymphocytes and Elimination of Cell-Associated Microbes

Naive T lymphocytes are activated by antigen and costimulators in peripheral lymphoid organs, and proliferate and differentiate into effector cells, most of which migrate to any site where the antigen (microbe) is present (Fig. 4–4). Upon activation, T lymphocytes secrete soluble proteins called *cytokines,* which function as growth and

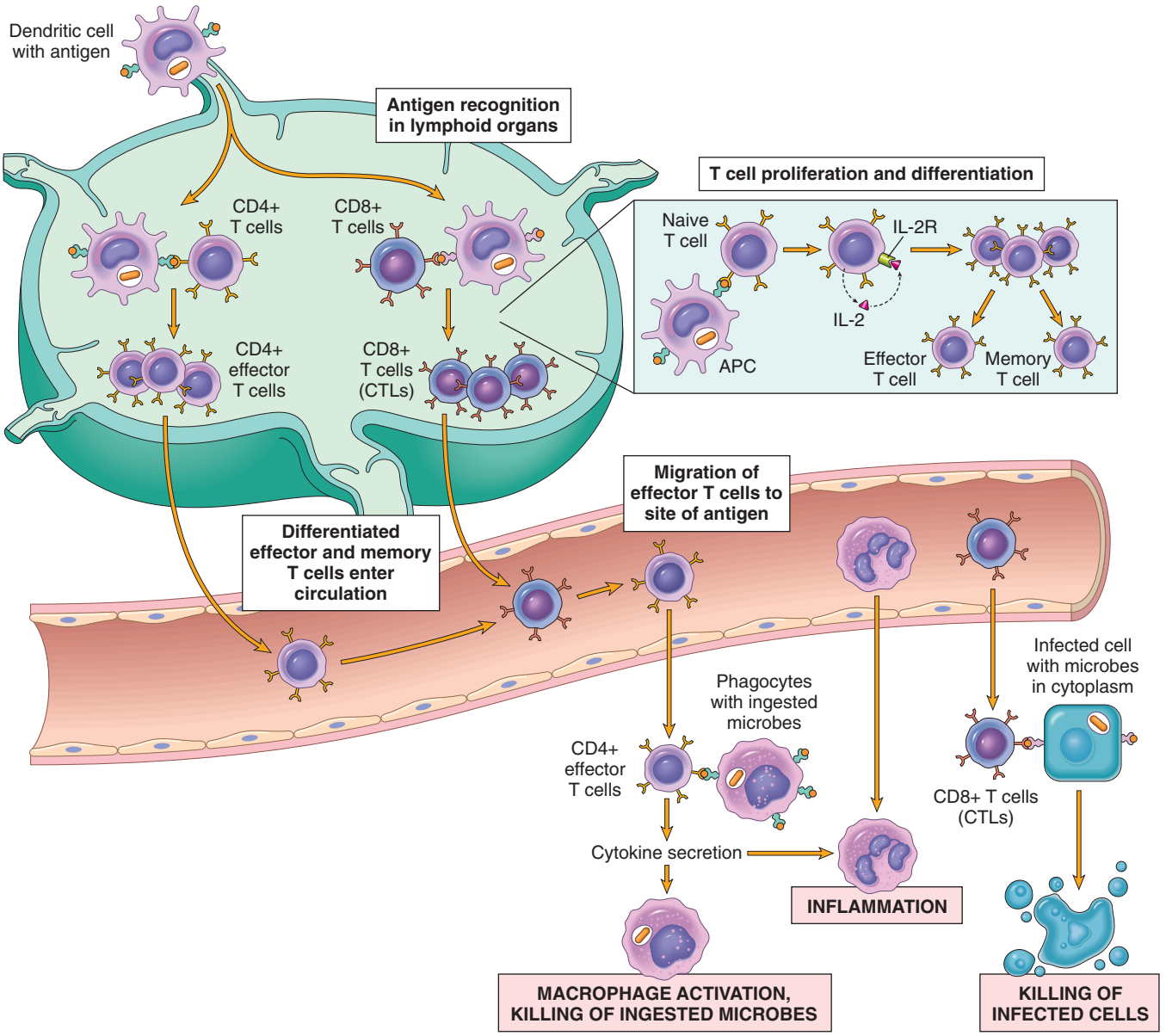


Figure 4–4 Cell-mediated immunity. Naive T cells recognize MHC-associated peptide antigens displayed on dendritic cells in lymph nodes. The T cells are activated to proliferate (under the influence of the cytokine IL-2) and to differentiate into effector and memory cells, which migrate to sites of infection and serve various functions in cell-mediated immunity. Effector CD4+ T cells of the T_H1 subset recognize the antigens of microbes ingested by phagocytes and activate the phagocytes to kill the microbes; T_H17 effector cells enhance leukocyte recruitment and stimulate inflammation; T_H2 cells activate eosinophils. CD8+ CTLs kill infected cells harboring microbes in the cytoplasm. Some activated T cells differentiate into long-lived memory cells. APC, antigen-presenting cell; CTLs, cytotoxic T lymphocytes.

differentiation factors for lymphocytes and other cells, and mediate communications between leukocytes. Because of the important roles of cytokines in both beneficial immune responses and in inflammatory diseases, it is important to understand their properties and actions.

Cytokines: Messenger Molecules of the Immune System

Cytokines are polypeptide products of many cell types (but principally activated lymphocytes and macrophages) that function as mediators of inflammation and immune responses. They were introduced in Chapter 2 in the context of inflammation; here we review their general properties and focus on those cytokines specifically involved in immunity.

Although different cytokines have diverse actions and functions, they all share some common features. Cytokines are synthesized and secreted in response to external stimuli, which may be microbial products, antigen recognition, or other cytokines. Their secretion typically is transient and is controlled by transcription and post-translational mechanisms. The actions of cytokines may be *autocrine* (on the cell that produces the cytokine), *paracrine* (on adjacent cells), and, less commonly, *endocrine* (at a distance from the site

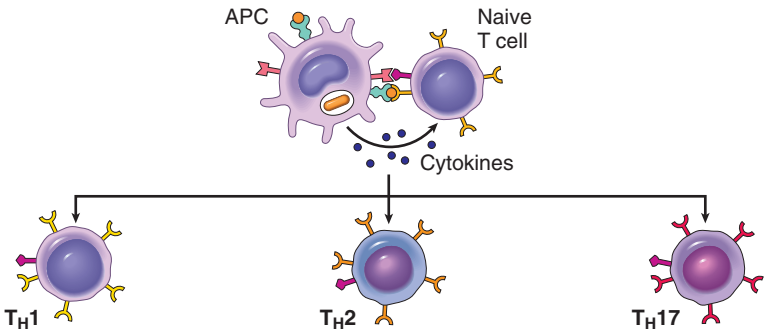
of production) (Chapter 2). The effects of cytokines tend to be pleiotropic (one cytokine can have diverse biologic activities, often on many cell types) and redundant (multiple cytokines may have the same activity). Molecularly defined cytokines are called interleukins, referring to their ability to mediate communications between leukocytes.

Cytokines may be grouped into several classes on the basis of their biologic activities and functions.

- *Cytokines involved in innate immunity and inflammation,* the earliest host response to microbes and dead cells. The major cytokines in this group are TNF and interleukin-1 (IL-1) and a group of chemoattractant cytokines called chemokines. IL-12, IFN-γ, IL-6, IL-23, and several other cytokines also participate in the early innate immune response. Major sources of these cytokines are activated macrophages and DCs, as well as endothelial cells, lymphocytes, mast cells, and other cell types. These were described in Chapter 2.
- *Cytokines that regulate lymphocyte responses and effector functions in adaptive immunity.* Different cytokines are involved in the proliferation and differentiation of lymphocytes (e.g., IL-2, IL-4), and in the activation of various effector cells (e.g., IFN-γ, which activates macrophages; IL-5, which activates eosinophils). The major sources of these cytokines are CD4+ helper T lymphocytes stimulated by antigens and costimulators. These cytokines are key participants in the induction and effector phases of adaptive cell-mediated immune responses (see later).
- *Cytokines that stimulate hematopoiesis.* Many of these are called colony-stimulating factors. They function to increase the output of leukocytes from the bone marrow and to thus replenish leukocytes that are consumed during immune and inflammatory reactions.

Effector Functions of T Lymphocytes

One of the earliest responses of CD4+ helper T cells is secretion of the cytokine IL-2 and expression of high-affinity receptors for IL-2. IL-2 is a growth factor that acts on these T lymphocytes and stimulates their proliferation, leading to an increase in the number of antigen-specific lymphocytes. Some of the progeny of the expanded pool of T cells differentiate into effector cells that can secrete different sets of cytokines and thus perform different functions. *The best-defined subsets of CD4+ helper cells are the T_H1, T_H2, and T_H17 subsets* (Fig. 4–5). T_H1 cells produce the cytokine IFN-γ, which activates macrophages and stimulates B cells to produce antibodies that activate complement and coat microbes for phagocytosis. T_H2 cells produce IL-4, which stimulates B cells to differentiate into IgE-secreting plasma cells; IL-5, which activates eosinophils; and IL-13, which activates mucosal epithelial cells to secrete mucus and expel microbes, and activates macrophages to secrete growth factors important for tissue repair. T_H17 cells produce the cytokine IL-17, which recruits neutrophils and thus promotes inflammation; T_H17 cells play an important role in some T cell–mediated inflammatory disorders. These effector cells migrate to sites of infection and accompanying tissue damage. When the differentiated effectors again encounter cell-associated microbes, they are activated to perform the functions that are responsible for elimination of the microbes. The key mediators of the functions of helper T cells are various cytokines and the surface molecule called CD40 ligand (CD40L), which binds to its receptor, CD40, on B cells and macrophages. Differentiated CD4+ effector T cells of the T_H1 subset recognize microbial peptides on macrophages that have ingested the microbes. The T cells express CD40L, which engages CD40 on the macrophages, and the T cells secrete the cytokine IFN-γ,



	T_H1	T_H2	T_H17
Cytokines produced	IFN-γ	IL-4, IL-5, IL-13	IL-17, IL-22, chemokines
Cytokines that induce this subset	IFN-γ, IL-12	IL-4	TGF-β, IL-6, IL-1, IL-23
Immunologic reactions triggered	Macrophage activation, stimulation of IgG antibody production	Stimulation of IgE production, activation of mast cells and eosinophils	Recruitment of neutrophils, monocytes
Host defense against	Intracellular microbes	Helminthic parasites	Extracellular bacteria, fungi
Role in disease	Immune-mediated chronic inflammatory diseases (often autoimmune)	Allergies	Immune-mediated chronic inflammatory diseases (often autoimmune)

Figure 4–5 Subsets of CD4+ effector T cells. In response to stimuli (mainly cytokines) present at the time of antigen recognition, naive CD4+ helper T cells may differentiate into populations of effector cells that produce distinct sets of cytokines and perform different functions. The types of immune reactions elicited by each subset, and its role in host defense and immunological diseases, are summarized. Two other populations of CD4+ T cells, regulatory cells and follicular helper cells, are not shown.

which is a potent macrophage activator. The combination of CD40- and IFN-γ–mediated activation results in the induction of potent microbicidal substances in the macrophages, including reactive oxygen species and nitric oxide, leading to the destruction of ingested microbes. T_H2 cells elicit cellular defense reactions that are dominated by eosinophils and not macrophages. As discussed later, CD4+ helper T cells also stimulate B cell responses by CD40L and cytokines. Some CD4+ T cells remain in the lymphoid organs in which they were activated and then migrate into follicles, where they stimulate antibody responses; these cells are called follicular helper T cells.

Activated CD8+ lymphocytes differentiate into CTLs, which kill cells harboring microbes in the cytoplasm. These microbes may be viruses that infect many cell types, or bacteria that are ingested by macrophages but have learned to escape from phagocytic vesicles into the cytoplasm (where they are inaccessible to the killing machinery of phagocytes, which is largely confined to vesicles). By destroying the infected cells, CTLs eliminate the reservoirs of infection.

Humoral Immunity: Activation of B Lymphocytes and Elimination of Extracellular Microbes

Upon activation, B lymphocytes proliferate and then differentiate into plasma cells that secrete different classes of antibodies with distinct functions (Fig. 4–6). There are two major mechanisms of B cell activation.

- *T cell–independent.* Many polysaccharide and lipid antigens have multiple identical antigenic determinants (epitopes) that are able to engage several antigen receptor molecules on each B cell and initiate the process of B cell activation.
- *T cell–dependent.* Typical globular protein antigens are not able to bind to many antigen receptors, and the full response of B cells to protein antigens requires help from CD4+ T cells. B cells also can act as APCs—they ingest protein antigens, degrade them, and display peptides bound to class II MHC molecules for recognition by helper T cells. The helper T cells express CD40L and secrete cytokines, which work together to activate the B cells.

Some of the progeny of the expanded B cell clones differentiate into antibody-secreting *plasma cells.* Each plasma cell secretes antibodies that have the same specificity as the cell surface antibodies (B cell receptors) that first recognized the antigen. Polysaccharides and lipids stimulate secretion mainly of IgM antibody. Protein antigens, by virtue of CD40L- and cytokine-mediated helper T cell actions, induce the production of antibodies of different classes (IgG, IgA, IgE). This production of functionally different antibodies, all with the same specificity, is called heavy-chain class (isotype) switching; it provides plasticity in the antibody response, allowing antibodies to serve many functions. Helper T cells also stimulate the production of antibodies with higher and higher affinity for the antigen. This process, called affinity maturation, improves the quality of the humoral immune response.

The humoral immune response combats microbes in numerous ways (Fig. 4–6).

- Antibodies bind to microbes and prevent them from infecting cells, thereby "neutralizing" the microbes.

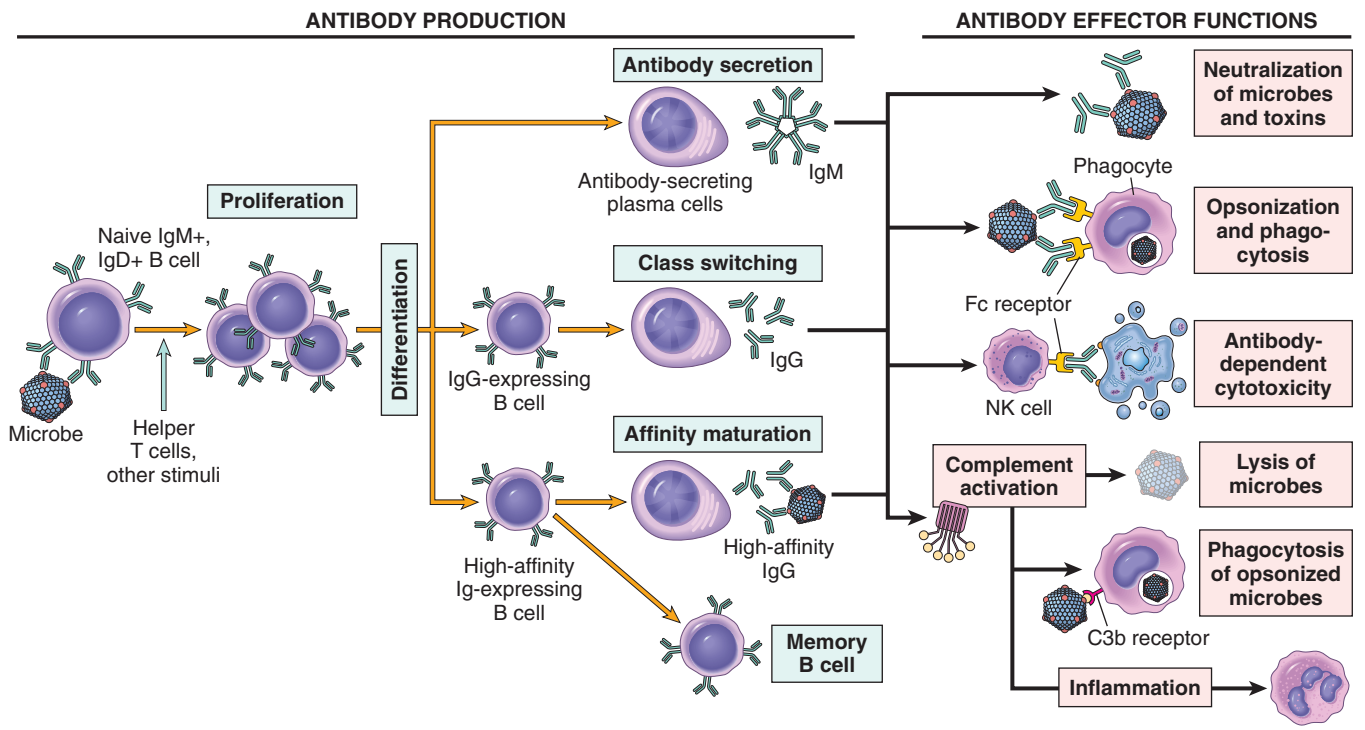


Figure 4–6 Humoral immunity. Naive B lymphocytes recognize antigens, and under the influence of helper T cells and other stimuli (not shown), the B cells are activated to proliferate and to differentiate into antibody-secreting plasma cells. Some of the activated B cells undergo heavy chain class switching and affinity maturation, and some become long-lived memory cells. Antibodies of different heavy chain isotypes (classes) perform different effector functions, shown on the right.

- IgG antibodies coat ("opsonize") microbes and target them for phagocytosis, since phagocytes (neutrophils and macrophages) express receptors for the Fc tails of IgG molecules.
- IgG and IgM activate the complement system by the classical pathway, and complement products promote phagocytosis and destruction of microbes. Production of most opsonizing and complement-fixing IgG antibodies is stimulated by IFN-γ, typically produced by T_H1 helper cells, which respond to many bacteria and viruses, and IgG antibodies are important mechanisms of defense against these microbes.
- IgA is secreted in mucosal tissues and neutralizes microbes in the lumens of the respiratory and gastrointestinal tracts (and other mucosal tissues).
- IgG is actively transported across the placenta and protects the newborn until the immune system becomes mature. This is called *passive immunity*.
- IgE coats helminthic parasites and functions with mast cells and eosinophils to kill them. As mentioned earlier, T_H2 helper cells secrete cytokines that stimulate the production of IgE and activate eosinophils, and thus the response to helminths is orchestrated by T_H2 cells.

Circulating IgG antibodies have half-lives of about 3 weeks, which is much longer than the half-lives of most blood proteins, as a consequence of special mechanisms for recycling IgG and reducing its catabolism. Some antibody-secreting plasma cells migrate to the bone marrow and live for years, continuing to produce low levels of antibodies.

Decline of Immune Responses and Immunologic Memory

A majority of effector lymphocytes induced by an infectious pathogen die by apoptosis after the microbe is eliminated, thus returning the immune system to its basal resting state. This return to a stable or steady state, called homeostasis, occurs because microbes provide essential stimuli for lymphocyte survival and activation, and effector cells are short-lived. Therefore, as the stimuli are eliminated, the activated lymphocytes are no longer kept alive.

The initial activation of lymphocytes also generates long-lived *memory cells*, which may survive for years after the infection. Memory cells are an expanded pool of antigen-specific lymphocytes (more numerous than the naive cells specific for any antigen that are present before encounter with that antigen), and memory cells respond faster and more effectively against the antigen than do naive cells. This is why the generation of memory cells is an important goal of vaccination.

This brief discussion of the normal immune response sets the stage for a consideration of the situations in which immune responses become abnormal, and of how these abnormalities lead to tissue injury and disease.

SUMMARY

Overview of Normal Immune Responses

- The physiologic function of the immune system is defense against infectious microbes.
- The early reaction to microbes is mediated by the mechanisms of innate immunity, which are ready to respond to microbes. These mechanisms include epithelial barriers, phagocytes, NK cells, and plasma proteins (e.g., of the complement system). The reaction of innate immunity is often manifested as inflammation.
- The defense reactions of adaptive immunity develop slowly, but are more potent and specialized.
- Microbes and other foreign antigens are captured by DCs and transported to lymph nodes, where the antigens are recognized by naive lymphocytes. The lymphocytes are activated to proliferate and differentiate into effector and memory cells.
- Cell-mediated immunity is the reaction of T lymphocytes, designed to combat cell-associated microbes (e.g., phagocytosed microbes and microbes in the cytoplasm of infected cells). Humoral immunity is mediated by antibodies and is effective against extracellular microbes (in the circulation and mucosal lumens).
- CD4+ helper T cells help B cells to make antibodies, activate macrophages to destroy ingested microbes, stimulate recruitment of leukocytes, and regulate all immune responses to protein antigens. The functions of CD4+ T cells are mediated by secreted proteins called cytokines. CD8+ CTLs kill cells that express antigens in the cytoplasm that are seen as foreign (e.g., virus-infected and tumor cells).
- Antibodies secreted by plasma cells neutralize microbes and block their infectivity, and promote the phagocytosis and destruction of pathogens. Antibodies also confer passive immunity to neonates.

HYPERSENSITIVITY REACTIONS: MECHANISMS OF IMMUNE-MEDIATED INJURY

Immune responses that normally are protective are also capable of causing tissue injury. Injurious immune reactions are grouped under *hypersensitivity*, and the resulting diseases are called hypersensitivity diseases. This term originated from the idea that persons who mount immune responses against an antigen are "sensitized" to that antigen, so pathologic or excessive reactions represent manifestations of a "hypersensitive" state. Normally, an exquisite system of checks and balances optimizes the eradication of infecting organisms without serious injury to host tissues. However, immune responses may be inadequately controlled or inappropriately targeted to host tissues, and in such situations, the normally beneficial response is the cause of disease. In this section we describe the causes and general mechanisms of hypersensitivity diseases and then discuss specific situations in which the immune response is responsible for the disease.

Causes of Hypersensitivity Reactions

Pathologic immune responses may be directed against different types of antigens and may result from various underlying abnormalities.

- *Autoimmunity: reactions against self antigens.* Normally, the immune system does not react against self-generated antigens. This phenomenon is called self tolerance, implying that the body "tolerates" its own antigens. On occasion, self-tolerance fails, resulting in reactions against the body's own cells and tissues; collectively, such reactions constitute autoimmunity. The diseases caused by autoimmunity are referred to as autoimmune diseases. We shall return to the mechanisms of self-tolerance and autoimmunity later in this chapter.
- *Reactions against microbes.* There are many types of reactions against microbial antigens that may cause disease. In some cases, the reaction appears to be excessive or the microbial antigen is unusually persistent. If antibodies are produced against such antigens, the antibodies may bind to the microbial antigens to produce immune complexes, which deposit in tissues and trigger inflammation; this is the underlying mechanism of poststreptococcal glomerulonephritis (Chapter 13). T cell responses against persistent microbes may give rise to severe inflammation, sometimes with the formation of granulomas (Chapter 2); this is the cause of tissue injury in tuberculosis and other infections. Rarely, antibodies or T cells reactive with a microbe cross-react with a host tissue; such cross-reactivity is believed to be the basis for rheumatic heart disease (Chapter 10). In some instances, the disease-causing immune response may be entirely normal, but in the process of eradicating the infection, host tissues are injured. In viral hepatitis, the virus that infects liver cells is not cytopathic, but it is recognized as foreign by the immune system. Cytotoxic T cells try to eliminate infected cells, and this normal immune response damages liver cells.
- *Reactions against environmental antigens.* Most healthy people do not react strongly against common environmental substances (e.g., pollens, animal danders, or dust mites), but almost 20% of the population are "allergic" to these substances. These individuals are genetically predisposed to make unusual immune responses to a variety of noninfectious, and otherwise harmless, antigens to which all persons are exposed but against which only some react.

In all of these conditions, tissue injury is caused by the same mechanisms that normally function to eliminate infectious pathogens—namely, antibodies, effector T lymphocytes, and various other effector cells. The problem in these diseases is that the response is triggered and maintained inappropriately. Because the stimuli for these abnormal immune responses are difficult or impossible to eliminate (e.g., self antigens, persistent microbes, or environmental antigens), and the immune system has many intrinsic positive feedback loops (amplification mechanisms), once a pathologic immune response starts it is difficult to control or terminate it. Therefore, these hypersensitivity diseases tend to be chronic and debilitating, and are therapeutic challenges. Since inflammation, typically chronic inflammation, is a major component of the pathology of these disorders, they are sometimes grouped under the rubric *immune-mediated inflammatory diseases.*

Types of Hypersensitivity Reactions

Hypersensitivity reactions are traditionally subdivided into four types based on the principal immune mechanism responsible for injury; three are variations on antibody-mediated injury, whereas the fourth is T cell–mediated (Table 4–1). The rationale for this classification is that the mechanism of immune injury is often a good predictor of the clinical manifestations and may even help to guide the therapy. However, this classification of immune-mediated diseases is not perfect, because several immune reactions may coexist in one disease.

- *Immediate (type I) hypersensitivity,* often called *allergy,* results from the activation of the T_H2 subset of CD4+ helper T cells by environmental antigens, leading to the production of IgE antibodies, which become attached to mast cells. When these IgE molecules bind the antigen (allergen), the mast cells are triggered to release mediators that transiently affect vascular permeability and

Table 4–1 Mechanisms of Hypersensitivity Reactions

Type	Immune Mechanisms	Histopathologic Lesions	Prototypical Disorders
Immediate (type I) hypersensitivity	Production of IgE antibody → immediate release of vasoactive amines and other mediators from mast cells; later recruitment of inflammatory cells	Vascular dilation, edema, smooth muscle contraction, mucus production, tissue injury, inflammation	Anaphylaxis; allergies; bronchial asthma (atopic forms)
Antibody-mediated (type II) hypersensitivity	Production of IgG, IgM → binds to antigen on target cell or tissue → phagocytosis or lysis of target cell by activated complement or Fc receptors; recruitment of leukocytes	Phagocytosis and lysis of cells; inflammation; in some diseases, functional derangements without cell or tissue injury	Autoimmune hemolytic anemia; Goodpasture syndrome
Immune complex–mediated (type III) hypersensitivity	Deposition of antigen–antibody complexes → complement activation → recruitment of leukocytes by complement products and Fc receptors → release of enzymes and other toxic molecules	Inflammation, necrotizing vasculitis (fibrinoid necrosis)	Systemic lupus erythematosus; some forms of glomerulonephritis; serum sickness; Arthus reaction
Cell-mediated (type IV) hypersensitivity	Activated T lymphocytes → (1) release of cytokines, inflammation and macrophage activation; (2) T cell–mediated cytotoxicity	Perivascular cellular infiltrates; edema; granuloma formation; cell destruction	Contact dermatitis; multiple sclerosis; type 1 diabetes; tuberculosis

IgE, IgG, IgM, immunoglobulins E, G, M.

induce smooth muscle contraction in various organs, and that also may stimulate more prolonged inflammation (the late-phase reaction). These diseases are commonly called allergic, or atopic, disorders.

- *Antibody-mediated (type II) hypersensitivity disorders* are caused by antibodies that bind to fixed tissue or cell surface antigens, promoting phagocytosis and destruction of the coated cells or triggering pathologic inflammation in tissues.
- *Immune complex–mediated (type III) hypersensitivity disorders* are caused by antibodies binding to antigens to form complexes that circulate and deposit in vascular beds and stimulate inflammation, typically as a consequence of complement activation. Tissue injury in these diseases is the result of the inflammation.
- *T cell–mediated (type IV) hypersensitivity disorders* are caused mainly by immune responses in which T lymphocytes of the T_H1 and T_H17 subsets produce cytokines that induce inflammation and activate neutrophils and macrophages, which are responsible for tissue injury. CD8+ CTLs also may contribute to injury by directly killing host cells.

Immediate (Type I) Hypersensitivity

Immediate hypersensitivity is a tissue reaction that occurs rapidly (typically within minutes) after the interaction of antigen with IgE antibody that is bound to the surface of mast cells in a sensitized host. The reaction is initiated by entry of an antigen, which is called an allergen because it triggers allergy. Many allergens are environmental substances that are harmless for most persons on exposure. Some people apparently inherit genes that make them susceptible to allergies. This susceptibility is manifested by the propensity of such persons to mount strong T_H2 responses and, subsequently, to produce IgE antibody against the allergens. The IgE is central to the activation of the mast cells and release of mediators that are responsible for the clinical and pathologic manifestations of the reaction. Immediate hypersensitivity may occur as a local reaction that is merely annoying (e.g., seasonal rhinitis, or hay fever), severely debilitating (asthma), or even fatal (anaphylaxis).

Sequence of Events in Immediate Hypersensitivity Reactions

Most hypersensitivity reactions follow the same sequence of cellular responses (Fig. 4–7):

- *Activation of T_H2 cells and production of IgE antibody.* Allergens may be introduced by inhalation, ingestion, or injection. Variables that probably contribute to the strong T_H2 responses to allergens include the route of entry, dose, and chronicity of antigen exposure, and the genetic makeup of the host. It is not clear if allergenic substances also have unique structural properties that endow them with the ability to elicit T_H2 responses. *Immediate hypersensitivity is the prototypical T_H2-mediated reaction.* The T_H2 cells that are induced secrete several cytokines, including IL-4, IL-5, and IL-13, which are responsible for essentially all the reactions of immediate hypersensitivity. IL-4 stimulates B cells specific for the allergen to undergo heavy-chain class switching to IgE and to secrete this immunoglobulin isotype. IL-5 activates eosinophils that are recruited to the reaction, and IL-13 acts on epithelial cells and stimulates mucus secretion. T_H2 cells often are recruited to the site of allergic reactions in response to chemokines that are produced locally; among these chemokines is eotaxin, which also recruits eosinophils to the same site.
- *Sensitization of mast cells by IgE antibody.* Mast cells are derived from precursors in the bone marrow, are widely distributed in tissues, and often reside near blood vessels and nerves and in subepithelial locations. Mast cells

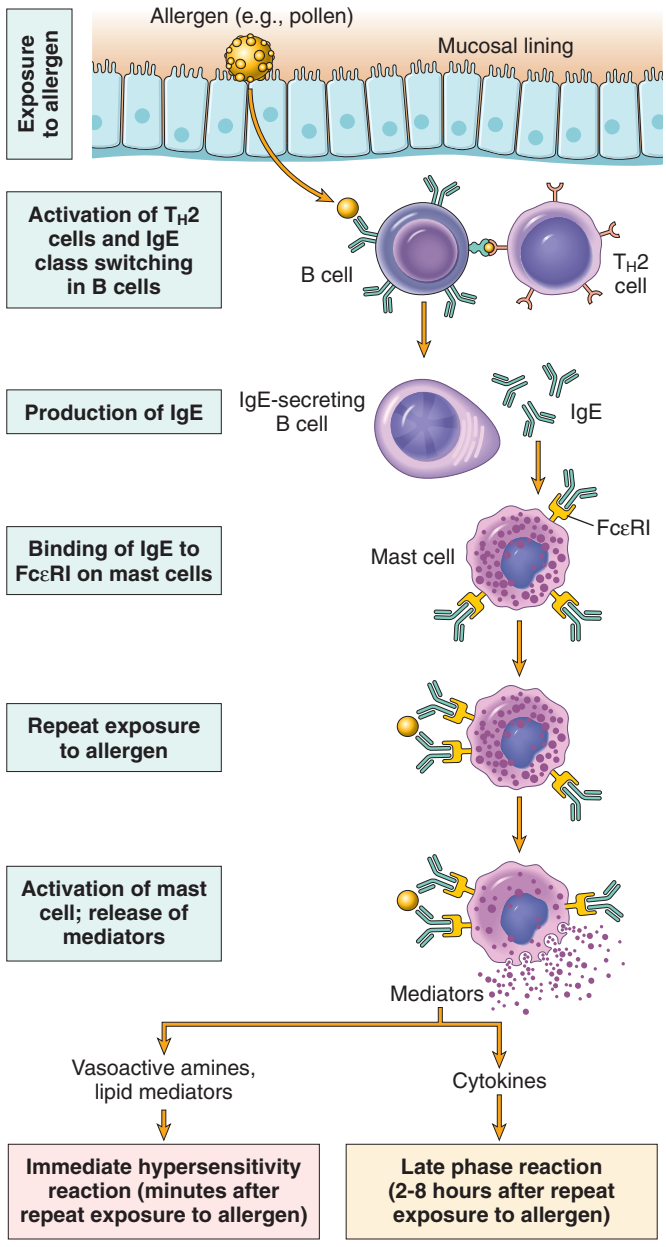


Figure 4–7 Sequence of events in immediate (type I) hypersensitivity. Immediate hypersensitivity reactions are initiated by the introduction of an allergen, which stimulates T_H2 responses and IgE production. IgE binds to Fc receptors (FcεRI) on mast cells, and subsequent exposure to the allergen activates the mast cells to secrete the mediators that are responsible for the pathologic manifestations of immediate hypersensitivity.

express a high-affinity receptor for the Fc portion of the ε heavy chain of IgE, called FcεRI. Even though the serum concentration of IgE is very low (in the range of 1 to 100 μg/mL), the affinity of the mast cell FcεRI receptor is so high that the receptors are always occupied by IgE. These antibody-bearing mast cells are "sensitized" to react if the antigen binds to the antibody molecules. Basophils are the circulating counterparts of mast cells. They also express FcεRI, but their role in most immediate hypersensitivity reactions is not established (since these reactions occur in tissues and not in the circulation). The third cell type that expresses FcεRI is eosinophils, which often are present in these reactions and also have a role in IgE-mediated host defense against helminth infections, described later.

- *Activation of mast cells and release of mediators.* When a person who was sensitized by exposure to an allergen is reexposed to the allergen, it binds to multiple specific IgE molecules on mast cells, usually at or near the site of allergen entry. When these IgE molecules are cross-linked, a series of biochemical signals is triggered in the mast cells. The signals culminate in the secretion of various mediators from the mast cells. Three groups of mediators are the most important in different immediate hypersensitivity reactions (Fig. 4–8):
  - *Vasoactive amines released from granule stores.* The granules of mast cells contain histamine, which is released within seconds or minutes of activation. Histamine causes vasodilation, increased vascular permeability, smooth muscle contraction, and increased secretion of mucus. Other rapidly released mediators include adenosine (which causes bronchoconstriction and inhibits platelet aggregation) and chemotactic factors for neutrophils and eosinophils. Other mast cell granule contents that may be secreted include several neutral proteases (e.g., tryptase), which may damage tissues and also generate kinins and cleave complement components to produce additional chemotactic and inflammatory factors (e.g., C3a) (Chapter 2). The granules also contain acidic proteoglycans (heparin, chondroitin sulfate), the main function of which seems to be as a storage matrix for the amines.
  - *Newly synthesized lipid mediators.* Mast cells synthesize and secrete prostaglandins and leukotrienes, by the same pathways as do other leukocytes (Chapter 2). These lipid mediators have several actions that are important in immediate hypersensitivity reactions. Prostaglandin D_2 (PGD_2) is the most abundant mediator generated by the cyclooxygenase pathway in mast cells. It causes intense bronchospasm as well as increased mucus secretion. The leukotrienes LTC_4 and LTD_4 are the most potent vasoactive and spasmogenic agents known; on a molar basis, they are several thousand times more active than histamine in increasing vascular permeability and in causing bronchial smooth muscle contraction. LTB_4 is highly chemotactic for neutrophils, eosinophils, and monocytes.
  - *Cytokines.* Activation of mast cells results in the synthesis and secretion of several cytokines that are important for the late-phase reaction. These include TNF and chemokines, which recruit and activate leukocytes (Chapter 2); IL-4 and IL-5, which amplify the T_H2-initiated immune reaction; and IL-13, which stimulates epithelial cell mucus secretion.

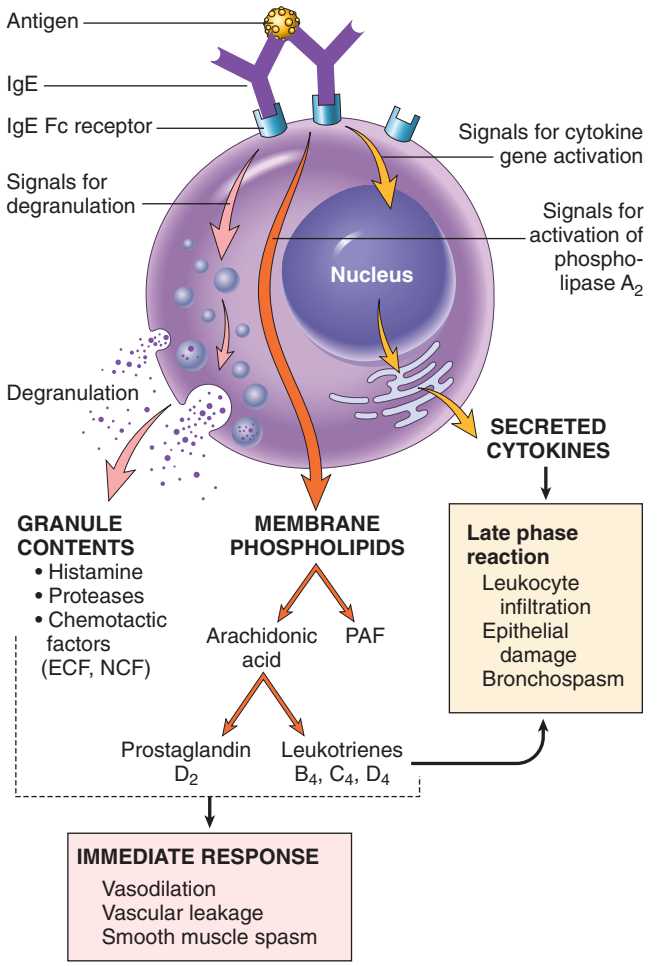


Figure 4–8 Mast cell mediators. Upon activation, mast cells release various classes of mediators that are responsible for the immediate and late-phase reactions. ECF, eosinophil chemotactic factor; NCF, neutrophil chemotactic factor (neither of these has been biochemically defined); PAF, platelet-activating factor.

In summary, a variety of compounds that act on blood vessels, smooth muscle, and leukocytes mediate type I hypersensitivity reactions (Table 4–2). Some of these compounds are released rapidly from sensitized mast cells and are responsible for the intense immediate reactions associated with conditions such as systemic anaphylaxis. Others, such as cytokines, are responsible for the inflammation seen in late-phase reactions.

Often, the IgE-triggered reaction has two well-defined phases (Fig. 4–9): (1) the *immediate response,* characterized by vasodilation, vascular leakage, and smooth muscle spasm, usually evident within 5 to 30 minutes after exposure to an allergen and subsiding by 60 minutes; and (2) a second, *late-phase reaction* that usually sets in 2 to 8 hours later and may last for several days and is characterized by inflammation as well as tissue destruction, such as mucosal epithelial cell damage. The dominant inflammatory cells in the late-phase reaction are neutrophils, eosinophils, and lymphocytes, especially T_H2 cells. Neutrophils are recruited by various chemokines; their roles in inflammation were described in Chapter 2. Eosinophils are recruited by eotaxin

Table 4–2 Summary of the Action of Mast Cell Mediators in Immediate (Type I) Hypersensitivity

Action	Mediators
Vasodilation, increased vascular permeability	Histamine PAF Leukotrienes C_4, D_4, E_4 Neutral proteases that activate complement and kinins Prostaglandin D_2
Smooth muscle spasm	Leukotrienes C_4, D_4, E_4 Histamine Prostaglandins PAF
Cellular infiltration	Cytokines (e.g., chemokines, TNF) Leukotriene B_4 Eosinophil and neutrophil chemotactic factors (not defined biochemically)

PAF, platelet-activating factor; TNF, tumor necrosis factor.

and other chemokines released from TNF-activated epithelium and are important effectors of tissue injury in the late-phase response. Eosinophils produce major basic protein and eosinophil cationic protein, which are toxic to epithelial cells, and LTC_4 and platelet-activating factor, which promote inflammation. T_H2 cells produce cytokines that have multiple actions, as described earlier. These recruited leukocytes can amplify and sustain the inflammatory response even in the absence of continuous allergen exposure. In addition, inflammatory leukocytes are responsible for much of the epithelial cell injury in immediate hypersensitivity. Because inflammation is a major component of many allergic diseases, notably asthma and atopic dermatitis, therapy usually includes anti-inflammatory drugs such as corticosteroids.

Clinical and Pathologic Manifestations

An immediate hypersensitivity reaction may occur as a systemic disorder or as a local reaction. The nature of the reaction is often determined by the route of antigen exposure. Systemic exposure to protein antigens (e.g., in bee venom) or drugs (e.g., penicillin) may result in systemic anaphylaxis. Within minutes of the exposure in a sensitized host, itching, urticaria (hives), and skin erythema appear, followed in short order by profound respiratory difficulty caused by pulmonary bronchoconstriction and accentuated by hypersecretion of mucus. Laryngeal edema may exacerbate matters by causing upper airway obstruction. In addition, the musculature of the entire gastrointestinal tract may be affected, with resultant vomiting, abdominal cramps, and diarrhea. Without immediate intervention, there may be systemic vasodilation with a fall in blood pressure (anaphylactic shock), and the patient may progress to circulatory collapse and death within minutes.

Local reactions generally occur when the antigen is confined to a particular site, such as skin (contact, causing urticaria), gastrointestinal tract (ingestion, causing diarrhea), or lung (inhalation, causing bronchoconstriction). The common forms of skin and food allergies, hay fever, and certain forms of asthma are examples of localized allergic reactions. However, ingestion or inhalation of allergens also can trigger systemic reactions.

Susceptibility to localized type I reactions has a strong genetic component, and the term *atopy* is used to imply familial predisposition to such localized reactions. Patients who suffer from nasobronchial allergy (including hay fever and some forms of asthma) often have a family history of similar conditions. Genes that are implicated in susceptibility to asthma and other atopic disorders include those encoding HLA molecules (which may confer immune responsiveness to particular allergens), cytokines (which may control T_H2 responses), a component of the FcεRI, and ADAM33, a metalloproteinase that may be involved in tissue remodeling in the airways.

The reactions of immediate hypersensitivity clearly did not evolve solely to cause human discomfort and disease. The immune response dependent on T_H2 cells and IgE—in particular, the late-phase inflammatory reaction—plays an important protective role in combating parasitic infections.

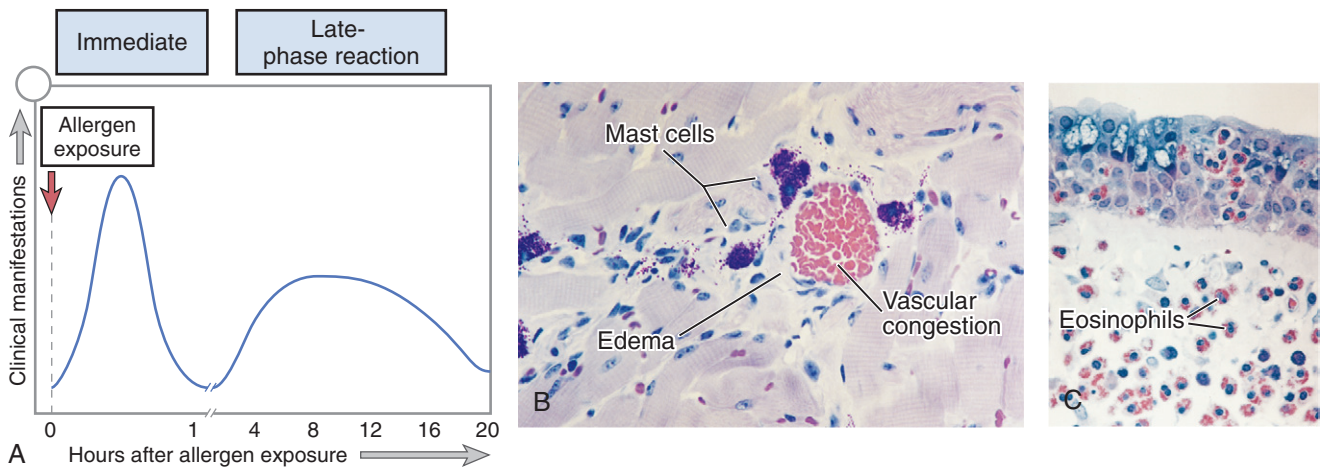


Figure 4–9 Immediate hypersensitivity. **A,** Kinetics of the immediate and late-phase reactions. The immediate vascular and smooth muscle reaction to allergen develops within minutes after challenge (allergen exposure in a previously sensitized person), and the late-phase reaction develops 2 to 24 hours later. **B–C,** Morphology: The immediate reaction **(B)** is characterized by vasodilation, congestion, and edema, and the late-phase reaction **(C)** is characterized by an inflammatory infiltrate rich in eosinophils, neutrophils, and T cells.

(B and C, Courtesy of Dr. Daniel Friend, Department of Pathology, Brigham and Women's Hospital, Boston, Massachusetts.)

IgE antibodies are produced in response to many helminthic infections, and their physiologic function is to target helminths for destruction by eosinophils and mast cells. Mast cells also are involved in defense against bacterial infections. And snake aficionados will be relieved to hear that their mast cells may protect them from some snake venoms by releasing granule proteases that degrade the toxins. Why these beneficial responses are inappropriately activated by harmless environmental antigens, giving rise to allergies, remains a puzzle.

SUMMARY

Immediate (Type I) Hypersensitivity

- Also called allergic reactions, or allergies
- Induced by environmental antigens (allergens) that stimulate strong T_H2 responses and IgE production in genetically susceptible individuals
- IgE coats mast cells by binding to Fcε receptors; reexposure to the allergen leads to cross-linking of the IgE and FcεRI, activation of mast cells, and release of mediators.
- Principal mediators are histamine, proteases, and other granule contents; prostaglandins and leukotrienes; and cytokines.
- Mediators are responsible for the immediate vascular and smooth muscle reactions and the late-phase reaction (inflammation).
- The clinical manifestations may be local or systemic, and range from mildly annoying rhinitis to fatal anaphylaxis.

Antibody-Mediated Diseases (Type II Hypersensitivity)

Antibody-mediated (type II) hypersensitivity disorders are caused by antibodies directed against target antigens on the surface of cells or other tissue components. The antigens may be normal molecules intrinsic to cell membranes or in the extracellular matrix, or they may be adsorbed exogenous antigens (e.g., a drug metabolite). Antibody-mediated abnormalities are the underlying cause of many human diseases; examples of these are listed in Table 4–3. In all of these disorders, the tissue damage or functional abnormalities result from a limited number of mechanisms.

Mechanisms of Antibody-Mediated Diseases

Antibodies cause disease by targeting cells for phagocytosis, by activating the complement system, and by interfering with normal cellular functions (Fig. 4–10). The antibodies that are responsible typically are high-affinity antibodies capable of activating complement and binding to the Fc receptors of phagocytes.

- *Opsonization and phagocytosis.* When circulating cells, such as erythrocytes or platelets, are coated (opsonized) with autoantibodies, with or without complement proteins, the cells become targets for phagocytosis by neutrophils and macrophages (Fig. 4–10, *A*). These phagocytes express receptors for the Fc tails of IgG antibodies and for breakdown products of the C3 complement protein, and use these receptors to bind and ingest opsonized particles. Opsonized cells are usually eliminated in the spleen, and this is why splenectomy is of

Table 4–3 Examples of Antibody-Mediated Diseases (Type II Hypersensitivity)

Disease	Target Antigen	Mechanisms of Disease	Clinicopathologic Manifestations
Autoimmune hemolytic anemia	Red cell membrane proteins (Rh blood group antigens, I antigen)	Opsonization and phagocytosis of erythrocytes	Hemolysis, anemia
Autoimmune thrombocytopenic purpura	Platelet membrane proteins (GpIIb/IIIa integrin)	Opsonization and phagocytosis of platelets	Bleeding
Pemphigus vulgaris	Proteins in intercellular junctions of epidermal cells (epidermal desmoglein)	Antibody-mediated activation of proteases, disruption of intercellular adhesions	Skin vesicles (bullae)
Vasculitis caused by ANCA	Neutrophil granule proteins, presumably released from activated neutrophils	Neutrophil degranulation and inflammation	Vasculitis
Goodpasture syndrome	Noncollagenous protein (NC1) in basement membranes of kidney glomeruli and lung alveoli	Complement- and Fc receptor–mediated inflammation	Nephritis, lung hemorrhage
Acute rheumatic fever	Streptococcal cell wall antigen; antibody cross-reacts with myocardial antigen	Inflammation, macrophage activation	Myocarditis
Myasthenia gravis	Acetylcholine receptor	Antibody inhibits acetylcholine binding, downmodulates receptors	Muscle weakness, paralysis
Graves disease (hyperthyroidism)	TSH receptor	Antibody-mediated stimulation of TSH receptors	Hyperthyroidism
Insulin-resistant diabetes	Insulin receptor	Antibody inhibits binding of insulin	Hyperglycemia, ketoacidosis
Pernicious anemia	Intrinsic factor of gastric parietal cells	Neutralization of intrinsic factor, decreased absorption of vitamin B_{12}	Abnormal myelopoiesis, anemia

ANCA, antineutrophil cytoplasmic antibodies; TSH, thyroid-stimulating hormone.

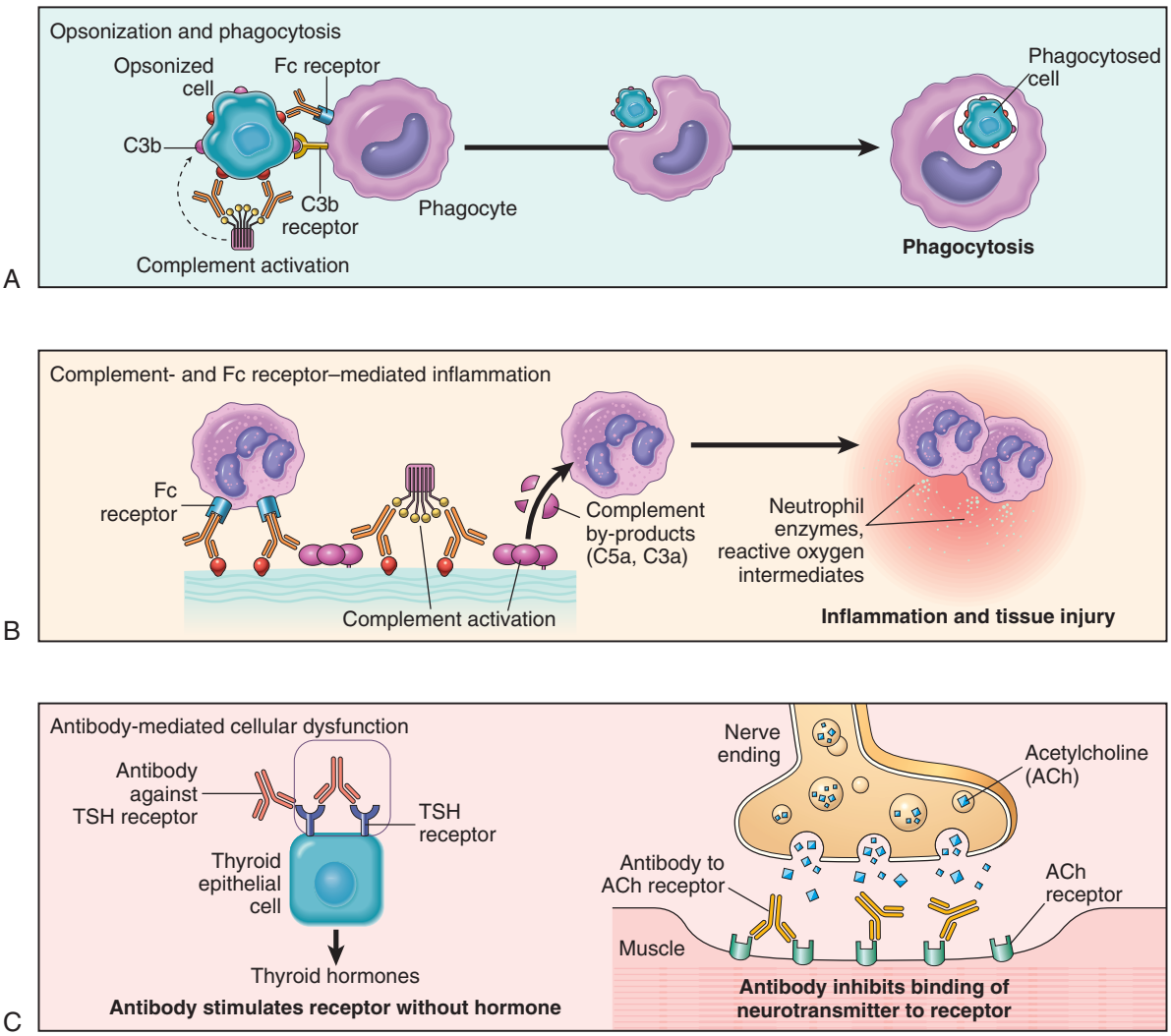


Figure 4–10 Mechanisms of antibody-mediated injury. **A,** Opsonization of cells by antibodies and complement components, and ingestion of opsonized cells by phagocytes. **B,** Inflammation induced by antibody binding to Fc receptors of leukocytes and by complement breakdown products. **C,** Antireceptor antibodies disturb the normal function of receptors. In these examples, antibodies against the thyroid-stimulating hormone (TSH) receptor activate thyroid cells in Graves disease, and acetylcholine (ACh) receptor antibodies impair neuromuscular transmission in myasthenia gravis.

clinical benefit in autoimmune thrombocytopenia and some forms of autoimmune hemolytic anemia.

- *Inflammation.* Antibodies bound to cellular or tissue antigens activate the complement system by the "classical" pathway (Fig. 4–10, *B*). Products of complement activation serve several functions (see Fig. 2–18, Chapter 2), one of which is to recruit neutrophils and monocytes, triggering inflammation in tissues. Leukocytes may also be activated by engagement of Fc receptors, which recognize the bound antibodies. This mechanism of injury is exemplified by Goodpasture syndrome and pemphigus vulgaris.
- *Antibody-mediated cellular dysfunction.* In some cases, antibodies directed against cell surface receptors impair or dysregulate cellular function without causing cell injury or inflammation (Fig. 4–10, *C*). In myasthenia gravis, antibodies against acetylcholine receptors in the motor end plates of skeletal muscles inhibit neuromuscular transmission, with resultant muscle weakness. Antibodies can also stimulate cellular responses excessively. In Graves disease, antibodies against the thyroid-stimulating hormone receptor stimulate thyroid epithelial cells to secrete thyroid hormones, resulting in hyperthyroidism. Antibodies against hormones and other essential proteins can neutralize and block the actions of these molecules, causing functional derangements.

Immune Complex Diseases (Type III Hypersensitivity)

Antigen–antibody (immune) complexes that are formed in the circulation may deposit in blood vessels, leading to complement activation and acute inflammation. The antigens in these complexes may be exogenous antigens, such as microbial proteins, or endogenous antigens, such as nucleoproteins. The mere formation of immune complexes does not equate with hypersensitivity disease; small amounts of antigen–antibody complexes may be produced during normal

Table 4–4 Examples of Immune Complex–Mediated Diseases

Disease	Antigen Involved	Clinicopathologic Manifestations
Systemic lupus erythematosus	Nuclear antigens	Nephritis, skin lesions, arthritis, others
Poststreptococcal glomerulonephritis	Streptococcal cell wall antigen(s); may be "planted" in glomerular basement membrane	Nephritis
Polyarteritis nodosa	Hepatitis B virus antigens in some cases	Systemic vasculitis
Reactive arthritis	Bacterial antigens (e.g., *Yersinia*)	Acute arthritis
Serum sickness	Various proteins (e.g., foreign serum protein such as horse antithymocyte globulin)	Arthritis, vasculitis, nephritis
Arthus reaction (experimental)	Various foreign proteins	Cutaneous vasculitis

immune responses and are usually phagocytosed and destroyed. It is only when these complexes are produced in large amounts, persist, and are deposited in tissues that they are pathogenic. Pathogenic immune complexes may form in the circulation and subsequently deposit in blood vessels, or the complexes may form at sites where antigen has been planted (in situ immune complexes). Immune complex–mediated injury is systemic when complexes are formed in the circulation and are deposited in several organs, or it may be localized to particular organs (e.g., kidneys, joints, or skin) if the complexes are formed and deposited in a specific site. The mechanism of tissue injury is the same regardless of the pattern of distribution; however, the sequence of events and the conditions leading to the formation of systemic and local immune complexes are different and are considered separately in the following descriptions. Immune complex diseases are some of the most common immunologic diseases (Table 4–4).

Systemic Immune Complex Disease

The pathogenesis of systemic immune complex disease can be divided into three phases: (1) formation of antigen–antibody complexes in the circulation and (2) deposition of the immune complexes in various tissues, thereby initiating (3) an inflammatory reaction in various sites throughout the body (Fig. 4–11).

Acute serum sickness is the prototype of a systemic immune complex disease. It was first described in humans when large amounts of foreign serum were administered for passive immunization (e.g., in persons receiving horse serum containing antidiphtheria antibody); it is now seen only rarely (e.g., in patients injected with rabbit or horse antithymocyte globulin for treatment of aplastic anemia or graft rejection, or patients with snakebite given anti-venom antibody made in animals). Although serum sickness is no longer common, the study of its pathogenesis sheds light on the mechanisms of human immune complex diseases. Approximately 5 days after the foreign protein is injected, specific antibodies are produced; these react with the antigen still present in the circulation to form antigen–antibody complexes. The complexes deposit in blood vessels in various tissue beds, triggering the subsequent injurious inflammatory reaction.

Several *variables determine whether immune complex formation leads to tissue deposition and disease*. Perhaps foremost among these factors is the size of the complexes. Very large complexes or complexes with many free IgG Fc regions (typically formed in antibody excess) are rapidly removed from the circulation by macrophages in the spleen and liver

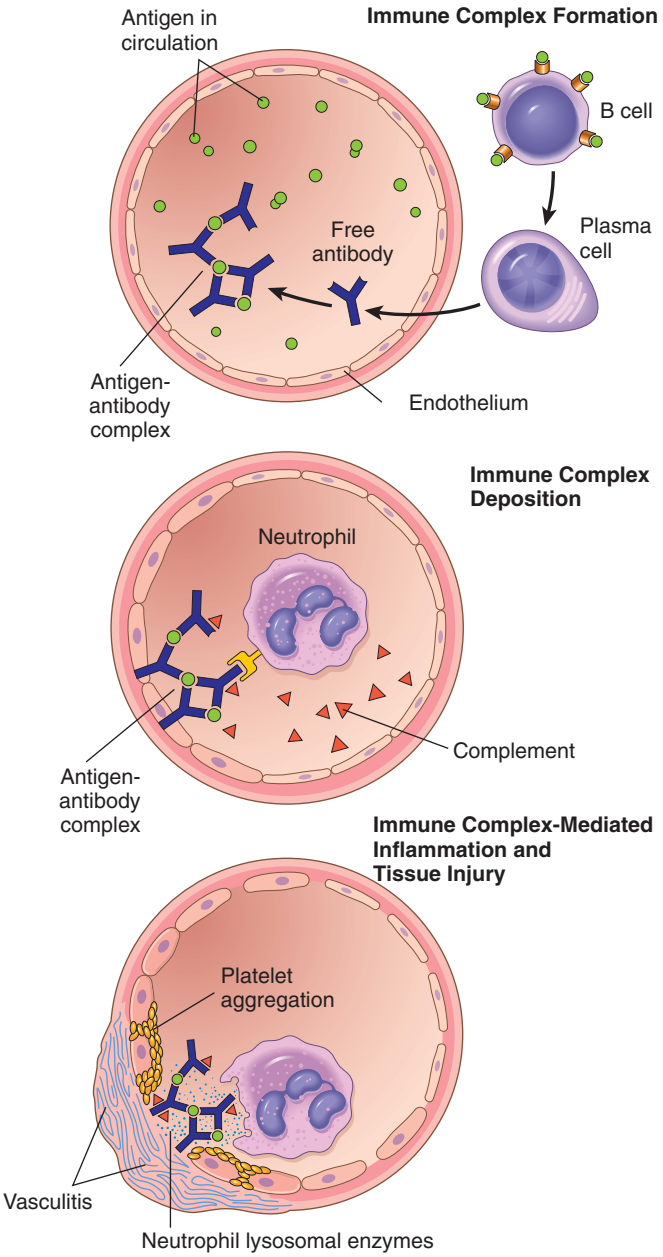


Figure 4–11 Immune complex disease: The sequential phases in the induction of systemic immune complex–mediated diseases (type III hypersensitivity).

and are therefore usually harmless. The most pathogenic complexes are formed during antigen excess and are small or intermediate in size and are cleared less effectively by phagocytes and therefore circulate longer. In addition, the charge of the complex, the valency of the antigen, the avidity of the antibody, and the hemodynamics of a given vascular bed all influence the tendency to develop disease. The favored sites of deposition are kidneys, joints, and small blood vessels in many tissues. Localization in the kidney and joints is explained in part by the high hemodynamic pressures associated with the filtration function of the glomerulus and the synovium. For complexes to leave the circulation and deposit within or outside the vessel wall, an increase in vascular permeability also must occur. This is probably triggered when immune complexes bind to leukocytes and mast cells by means of Fc and C3b receptors and stimulate release of mediators that increase vascular permeability.

Once complexes are deposited in the tissue, the third phase, the *inflammatory reaction*, ensues. During this phase (approximately 10 days after antigen administration), clinical features such as fever, urticaria, arthralgias, lymph node enlargement, and proteinuria appear. Wherever immune complexes deposit, characteristic tissue damage occurs. The immune complexes activate the complement system, leading to the release of biologically active fragments such as the anaphylatoxins (C3a and C5a), which increase vascular permeability and are chemotactic for neutrophils and monocytes (Chapter 2). The complexes also bind to Fcγ receptors on neutrophils and monocytes, activating these cells. Attempted phagocytosis of immune complexes by the leukocytes results in the secretion of a variety of additional pro-inflammatory substances, including prostaglandins, vasodilator peptides, and chemotactic substances, as well as lysosomal enzymes capable of digesting basement membrane, collagen, elastin, and cartilage, and reactive-oxygen species that damage tissues. Immune complexes can also cause platelet aggregation and activate Hageman factor; both of these reactions augment the inflammatory process and initiate formation of microthrombi, which contribute to the tissue injury by producing local ischemia (Fig. 4–11). The resultant pathologic lesion is termed *vasculitis* if it occurs in blood vessels, *glomerulonephritis* if it occurs in renal glomeruli, *arthritis* if it occurs in the joints, and so on.

Predictably, the antibody classes that induce such lesions are complement-fixing antibodies (i.e., IgG and IgM) and antibodies that bind to phagocyte Fc receptors (IgG). During the active phase of the disease, consumption of complement may result in decreased serum complement levels. The role of complement- and Fc receptor–dependent inflammation in the pathogenesis of the tissue injury is supported by the observation that experimental depletion of serum complement levels or knockout of Fc receptors in mice greatly reduces the severity of lesions, as does depletion of neutrophils.

MORPHOLOGY

The morphologic appearance of immune complex injury is dominated by acute **necrotizing vasculitis,** microthrombi, and superimposed ischemic necrosis accompanied by acute inflammation of the affected organs. The necrotic vessel wall takes on a smudgy eosinophilic appearance called **fibrinoid necrosis,** caused by protein deposition (see Fig. 1–13, Chapter 1). Immune complexes can be visualized in the tissues, usually in the vascular wall (examples of such deposits in the kidney in lupus are shown in Fig. 4–18, *E*). In due course, the lesions tend to resolve, especially when they were brought about by a single exposure to antigen (e.g., in acute serum sickness or acute poststreptococcal glomerulonephritis) (Chapter 13). However, chronic immune complex disease develops when there is persistent antigenemia or repeated exposure to an antigen. This occurs in some human diseases, such as systemic lupus erythematosus (SLE). Most often, even though the morphologic changes and other findings strongly implicate immune complex disease, the inciting antigens are unknown.

Local Immune Complex Disease

A model of local immune complex diseases is the *Arthus reaction,* in which an area of tissue necrosis appears as a result of acute immune complex vasculitis. The reaction is produced experimentally by injecting an antigen into the skin of a previously immunized animal (i.e., pre-formed antibodies against the antigen are already present in the circulation). Because of the initial antibody excess, immune complexes are formed as the antigen diffuses into the vascular wall; these are precipitated at the site of injection and trigger the same inflammatory reaction and histologic appearance as in systemic immune complex disease. Arthus lesions evolve over a few hours and reach a peak 4 to 10 hours after injection, when the injection site develops visible edema with severe hemorrhage, occasionally followed by ulceration.

SUMMARY

Pathogenesis of Diseases Caused by Antibodies and Immune Complexes

- Antibodies can coat (opsonize) cells, with or without complement proteins, and target these cells for phagocytosis by macrophages, which express receptors for the Fc tails of IgG molecules and for complement proteins. The result is depletion of the opsonized cells.
- Antibodies and immune complexes may deposit in tissues or blood vessels, and elicit an acute inflammatory reaction by activating complement, with release of breakdown products, or by engaging Fc receptors of leukocytes. The inflammatory reaction causes tissue injury.
- Antibodies can bind to cell surface receptors or essential molecules, and cause functional derangements (either inhibition or unregulated activation) without cell injury.

T Cell–Mediated (Type IV) Hypersensitivity

Several autoimmune disorders, as well as pathologic reactions to environmental chemicals and persistent microbes, are now known to be caused by T cells (Table 4–5). The occurrence and significance of T lymphocyte–mediated tissue injury

Table 4–5 T Cell–Mediated Diseases*

Disease	Specificity of Pathogenic T Cells	Principal Mechanisms of Tissue Injury	Clinicopathologic Manifestations
Rheumatoid arthritis	Collagen?; citrullinated self proteins?	Inflammation mediated by T_H17 (and T_H1?) cytokines; role of antibodies and immune complexes?	Chronic arthritis with inflammation, destruction of articular cartilage and bone
Multiple sclerosis	Protein antigens in myelin (e.g., myelin basic protein)	Inflammation mediated by T_H1 and T_H17 cytokines, myelin destruction by activated macrophages	Demyelination in CNS with perivascular inflammation; paralysis, ocular lesions
Type I diabetes mellitus	Antigens of pancreatic islet β cells (insulin, glutamic acid decarboxylase, others)	T cell–mediated inflammation, destruction of islet cells by CTLs	Insulitis (chronic inflammation in islets), destruction of β cells; diabetes
Hashimoto thyroiditis	Thyroglobulin, other thyroid proteins	Inflammation, CTL-mediated killing of thyroid epithelial cells	Hypothyroidism
Inflammatory bowel disease	Enteric bacteria; self antigens?	Inflammation mediated mainly by T_H17 cytokines	Chronic intestinal inflammation, ulceration, obstruction
Autoimmune myocarditis	Myosin heavy chain protein	CTL-mediated killing of myocardial cells; inflammation mediated by T_H1 cytokines	Cardiomyopathy
Contact sensitivity	Various environmental chemicals (e.g., urushiol from poison ivy or poison oak)	Inflammation mediated by T_H1 (and T_H17?) cytokines	Epidermal necrosis, dermal inflammation with skin rash and blisters

*Examples of human T cell–mediated diseases are listed. In many cases, the specificity of the T cells and the mechanisms of tissue injury are inferred on the basis of similarity to experimental animal models of the diseases.
CNS, central nervous system; CTL, cytotoxic T lymphocyte.

have been increasingly appreciated as the methods for detecting and purifying T cells from patients' circulation and lesions have improved. This group of diseases is of great clinical interest because many of the new, rationally designed biologic therapies for immune-mediated inflammatory diseases have been developed to target abnormal T cell reactions. Two types of T cell reactions are capable of causing tissue injury and disease: (1) cytokine-mediated inflammation, in which the cytokines are produced mainly by CD4+ T cells, and (2) direct cell cytotoxicity, mediated by CD8+ T cells (Fig. 4–12). In inflammation, exemplified by the delayed-type hypersensitivity (DTH) reaction, CD4+ T cells of the T_H1 and T_H17 subsets secrete cytokines, which recruit and activate other cells, especially macrophages, and these are the major effector cells of injury. In cell-mediated cytotoxicity, cytotoxic CD8+ T cells are responsible for tissue damage.

Inflammatory Reactions Elicited by CD4+ T Cells

The sequence of events in T cell–mediated inflammatory reactions begins with the first exposure to antigen and is essentially the same as the reactions of cell-mediated immunity (Fig. 4–4). Naive CD4+ T lymphocytes recognize peptide antigens of self or microbial proteins in association with class II MHC molecules on the surface of DCs (or macrophages) that have processed the antigens. If the DCs produce IL-12, the naive T cells differentiate into effector cells of the T_H1 type. The cytokine IFN-γ, made by NK cells and by the T_H1 cells themselves, further promotes T_H1 differentiation, providing a powerful positive feedback loop. If the APCs produce IL-1, IL-6, or IL-23 instead of IL-12, the CD4+ cells develop into T_H17 effectors. On subsequent exposure to the antigen, the previously generated effector cells are recruited to the site of antigen exposure and are activated by the antigen presented by local APCs. The T_H1 cells secrete IFN-γ, which is the most potent macrophage-activating cytokine known. Activated macrophages have increased phagocytic and microbicidal activity. Activated macrophages also express more class II MHC molecules and costimulators, leading to augmented antigen presentation capacity, and the cells secrete more IL-12, thus stimulating more T_H1 responses. Upon activation by antigen, T_H17 effector cells secrete IL-17 and several other cytokines, which promote the recruitment of neutrophils (and monocytes) and thus induce inflammation. Because the cytokines produced by the T cells enhance leukocyte recruitment and activation, these inflammatory reactions become chronic unless the offending agent is eliminated or the cycle is interrupted therapeutically. In fact, inflammation occurs as an early response to microbes and dead cells (Chapter 2), but it is greatly increased and prolonged when T cells are involved.

Delayed-type hypersensitivity (DTH), described next, is an illustrative model of T cell–mediated inflammation and tissue injury. The same reactions are the underlying basis for several diseases. *Contact dermatitis* is an example of tissue injury resulting from T cell–mediated inflammation. It is evoked by contact with pentadecylcatechol (also known as urushiol, the active component of poison ivy and poison oak, which probably becomes antigenic by binding to a host protein). On reexposure of a previously exposed person to the plants, sensitized T_H1 CD4+ cells accumulate in the dermis and migrate toward the antigen within the epidermis. Here they release cytokines that damage keratinocytes, causing separation of these cells and formation of an intraepidermal vesicle, and inflammation manifested as a vesicular dermatitis. It has long been thought that several *systemic diseases,* such as type 1 diabetes and multiple sclerosis, are caused by T_H1 and T_H17 reactions against self antigens, and Crohn disease may be caused by uncontrolled reactions involving the same T cells but directed against intestinal bacteria. T cell–mediated inflammation also plays a role in the rejection of transplants, described later in the chapter.

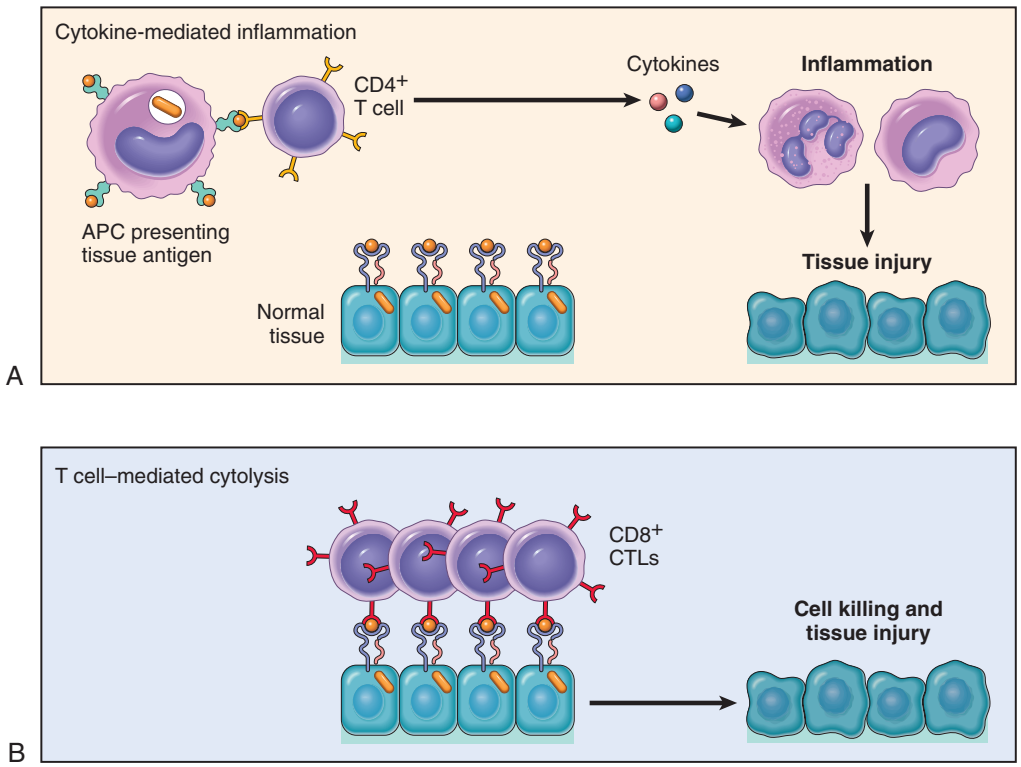


Figure 4–12 Mechanisms of T cell–mediated (type IV) hypersensitivity reactions. **A,** In cytokine-mediated inflammatory reactions, CD4+ T cells respond to tissue antigens by secreting cytokines that stimulate inflammation and activate phagocytes, leading to tissue injury. **B,** In some diseases, CD8+ CTLs directly kill tissue cells. APC, antigen-presenting cell; CTLs, cytotoxic T lymphocytes.

Delayed-Type Hypersensitivity

DTH is a T cell–mediated reaction that develops in response to antigen challenge in a previously sensitized individual. In contrast with immediate hypersensitivity, the DTH reaction is delayed for 12 to 48 hours, which is the time it takes for effector T cells to be recruited to the site of antigen challenge and to be activated to secrete cytokines. The classic example of DTH is the tuberculin reaction, elicited by challenge with a protein extract of *M. tuberculosis* (tuberculin) in a person who has previously been exposed to the tubercle bacillus. Between 8 and 12 hours after intracutaneous injection of tuberculin, a local area of erythema and induration appears, reaching a peak (typically 1 to 2 cm in diameter) in 24 to 72 hours and thereafter slowly subsiding. On histologic examination, the DTH reaction is characterized by perivascular accumulation ("cuffing") of CD4+ helper T cells and macrophages (Fig. 4–13). Local secretion of cytokines by these cells leads to increased microvascular permeability, giving rise to dermal edema and fibrin deposition; the latter is the main cause of the tissue induration in these responses. DTH reactions are mediated primarily by T_H1 cells; the contribution of T_H17 cells is unclear. The tuberculin response is used to screen populations for people who have had previous exposure to tuberculosis and therefore have circulating memory T cells specific for mycobacterial proteins. Notably, immunosuppression or loss of CD4+ T cells (e.g., resulting from HIV infection) may lead to a negative tuberculin response even in the presence of a severe infection.

Prolonged DTH reactions against persistent microbes or other stimuli may result in a special morphologic pattern of reaction called *granulomatous inflammation.* The initial perivascular CD4+ T cell infiltrate is progressively replaced by macrophages over a period of 2 to 3 weeks. These accumulated macrophages typically exhibit morphologic evidence of activation; that is, they become large, flat, and eosinophilic, and are called epithelioid cells. The epithelioid cells occasionally fuse under the influence of cytokines (e.g., IFN-γ) to form multinucleate giant cells. A microscopic aggregate of epithelioid cells, typically surrounded by a collar of lymphocytes, is called a granuloma (Fig. 4–14, *A*). The process is essentially a chronic form of T_H1-mediated inflammation and macrophage activation (Fig. 4–14, *B*). Older granulomas develop an enclosing rim of fibroblasts and connective tissue. Recognition of a granuloma is of diagnostic importance because of the limited number of conditions that can cause it (Chapter 2).

T Cell–Mediated Cytotoxicity

In this form of T cell–mediated tissue injury, CD8+ CTLs kill antigen-bearing target cells. As discussed earlier, class I MHC molecules bind to intracellular peptide antigens and present the peptides to CD8+ T lymphocytes, stimulating the differentiation of these T cells into effector cells called CTLs. CTLs play a critical role in resistance to virus infections and some tumors. The principal mechanism of killing by CTLs is dependent on the perforin–granzyme system. Perforin and granzymes are stored in the granules of CTLs and are rapidly released when CTLs engage their targets (cells bearing the appropriate class I MHC–bound peptides). Perforin binds to the plasma membrane of the

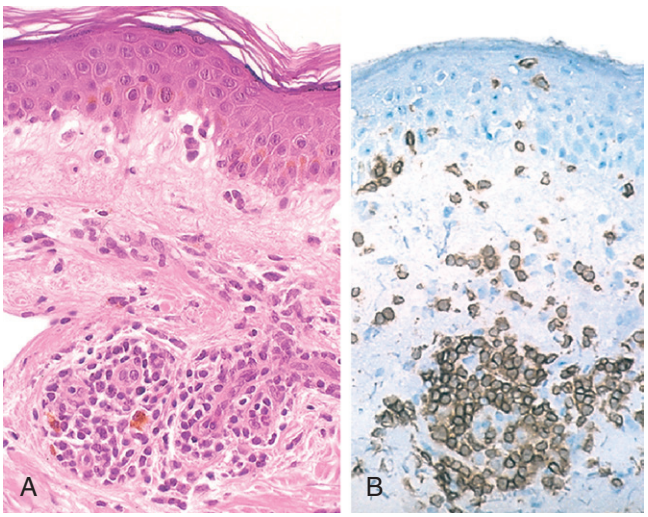


Figure 4–13 Delayed-type hypersensitivity reaction in the skin. **A,** Perivascular accumulation ("cuffing") of mononuclear inflammatory cells (lymphocytes and macrophages), with associated dermal edema and fibrin deposition. **B,** Immunoperoxidase staining reveals a predominantly perivascular cellular infiltrate that marks positively with anti-CD4 antibodies.

(B, Courtesy of Dr. Louis Picker, Department of Pathology, Oregon Health & Science University, Portland, Oregon.)

target cells and promotes the entry of granzymes, which are proteases that specifically cleave and thereby activate cellular caspases. These enzymes induce apoptotic death of the target cells (Chapter 1). CTLs play an important role in the rejection of solid-organ transplants and may contribute to many immunologic diseases, such as type 1 diabetes (in which insulin-producing β cells in pancreatic islets are destroyed by an autoimmune T cell reaction). CD8+ T cells may also secrete IFN-γ and contribute to cytokine-mediated inflammation, but less so than CD4+ cells.

SUMMARY

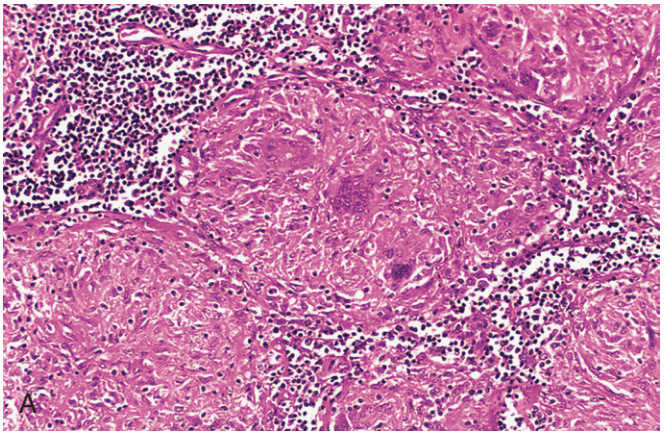
Mechanisms of T Cell–Mediated Hypersensitivity Reactions

- Cytokine-mediated inflammation: CD4+ T cells are activated by exposure to a protein antigen and differentiate into T_H1 and T_H17 effector cells. Subsequent exposure to the antigen results in the secretion of cytokines. IFN-γ activates macrophages to produce substances that cause tissue damage and promote fibrosis, and IL-17 and other cytokines recruit leukocytes, thus promoting inflammation.
- T cell–mediated cytotoxicity: CD8+ CTLs specific for an antigen recognize cells expressing the target antigen and kill these cells. CD8+ T cells also secrete IFN-γ.

With the basic mechanisms of pathologic immune reactions as background, we now proceed to a consideration of two categories of reactions that are of great clinical importance: autoimmunity and transplant rejection.

AUTOIMMUNE DISEASES

Immune reactions to self antigens (i.e., autoimmunity) are the underlying cause of numerous human diseases. Autoimmune diseases currently are estimated to affect 2% to 5% of the population in developed countries, and appear to be increasing in incidence. The evidence that these diseases are indeed the result of autoimmune reactions is more persuasive for some than for others. For instance, in many of these disorders, multiple high-affinity autoantibodies have



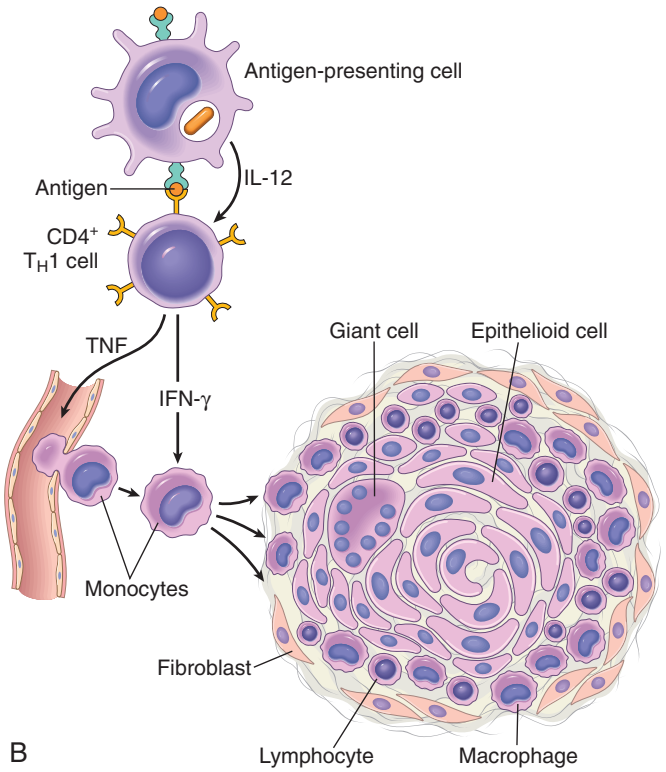


Figure 4–14 Granulomatous inflammation. **A,** A section of a lymph node shows several granulomas, each made up of an aggregate of epithelioid cells and surrounded by lymphocytes. The granuloma in the center shows several multinucleate giant cells. **B,** The events that give rise to the formation of granulomas in type IV hypersensitivity reactions. Note the role played by T cell–derived cytokines.

(A, Courtesy of Dr. Trace Worrell, Department of Pathology, University of Texas Southwestern Medical School, Dallas, Texas.)

Table 4–6 Autoimmune Diseases

Organ-Specific	Systemic
Diseases Mediated by Antibodies	
Autoimmune hemolytic anemia	Systemic lupus erythematosus
Autoimmune thrombocytopenia	
Autoimmune atrophic gastritis of pernicious anemia	
Myasthenia gravis	
Graves disease	
Goodpasture syndrome	
Diseases Mediated by T Cells*	
Type I diabetes mellitus	Rheumatoid arthritis
Multiple sclerosis	Systemic sclerosis (scleroderma)
Hashimoto thyroiditis	Sjögren syndrome
Crohn disease	
Diseases Postulated to Be of Autoimmune Origin†	
Primary biliary cirrhosis	Polyarteritis nodosa
Autoimmune (chronic active) hepatitis	Inflammatory myopathies

*A role for T cells has been demonstrated in these disorders, but antibodies also may be involved in tissue injury.
†An autoimmune basis for these disorders is suspected, but the supporting evidence is not strong.

been identified, and in some cases these antibodies are known to cause pathologic abnormalities (Table 4–6). Similarly, with improving technology, there is growing evidence for the activation of pathogenic self-reactive T cells in some of these diseases. In addition, experimental models have proved very informative, providing circumstantial evidence supporting an autoimmune etiology. Nevertheless, it is fair to say that for many disorders traditionally classified as autoimmune, this etiologic origin is suspected but not proved.

Presumed autoimmune diseases range from those in which specific immune responses are directed against one particular organ or cell type and result in localized tissue damage, to multisystem diseases characterized by lesions in many organs and associated with multiple autoantibodies or T cell–mediated reactions against numerous self antigens. In many of the systemic diseases that are caused by immune complexes and autoantibodies, the lesions affect principally the connective tissue and blood vessels of the various organs involved. Therefore, these diseases are often referred to as "collagen vascular" or "connective tissue" disorders, even though the immunologic reactions are not specifically directed against constituents of connective tissue or blood vessels.

Normal persons are unresponsive (tolerant) to their own (self) antigens, and autoimmunity results from a failure of self-tolerance. Therefore, understanding the pathogenesis of autoimmunity requires familiarity with the mechanisms of normal immunologic tolerance.

Immunologic Tolerance

Immunologic tolerance is unresponsiveness to an antigen that is induced by exposure of specific lymphocytes to that antigen. Self-tolerance refers to a lack of immune responsiveness to one's own tissue antigens. Billions of different antigen receptors are randomly generated in developing T and B lymphocytes, and it is not surprising that during this process, receptors are produced that can recognize self antigens. Since these antigens cannot all be concealed from the immune system, there must be means of eliminating or controlling self-reactive lymphocytes. Several mechanisms work in concert to select against self-reactivity and to thus prevent immune reactions against the body's own antigens. These mechanisms are broadly divided into two groups: central tolerance and peripheral tolerance (Fig. 4–15).

Central tolerance. The principal mechanism of central tolerance is the antigen-induced deletion (death) of self-reactive T and B lymphocytes during their maturation in central (generative) lymphoid organs (i.e., in the thymus for T cells and in the bone marrow for B cells). In the thymus, many autologous (self) protein antigens are processed and presented by thymic APCs in association with self MHC. Any immature T cell that encounters such a self antigen undergoes apoptosis (a process called deletion, or negative selection), and the T cells that complete their maturation are thereby depleted of self-reactive cells (Fig. 4–15). An exciting advance has been the identification of putative transcription factors that induce the expression of peripheral tissue antigens in the thymus, thus making the thymus an immunologic mirror of self. One such factor is called the autoimmune regulator (AIRE); mutations in the *AIRE* gene are responsible for an autoimmune polyendocrine syndrome in which T cells specific for multiple self antigens escape deletion (presumably because these self antigens are not expressed in the thymus), and attack tissues expressing the self antigens. Some T cells that encounter self antigens in the thymus are not killed but differentiate into regulatory T cells, as described later.

Immature B cells that recognize self antigens with high affinity in the bone marrow also may die by apoptosis. Some self-reactive B cells may not be deleted but may undergo a second round of rearrangement of antigen receptor genes and then express new receptors that are no longer self-reactive (a process called "receptor editing").

Unfortunately, the process of deletion of self-reactive lymphocytes is not perfect. Many self antigens may not be present in the thymus, so T cells bearing receptors for such autoantigens can escape into the periphery. There is similar "slippage" in the B cell system as well, and B cells that bear receptors for a variety of self antigens, including thyroglobulin, collagen, and DNA, can be found in healthy persons.

Peripheral tolerance. Self-reactive T cells that escape negative selection in the thymus can potentially wreak havoc unless they are deleted or effectively muzzled. Several mechanisms in the peripheral tissues that silence such potentially autoreactive T cells have been identified (Fig. 4–15):

- *Anergy*: This term refers to functional inactivation (rather than death) of lymphocytes induced by encounter with antigens under certain conditions. As described previously, activation of T cells requires two signals: recognition of peptide antigen in association with self MHC molecules on APCs, and a set of second costimulatory

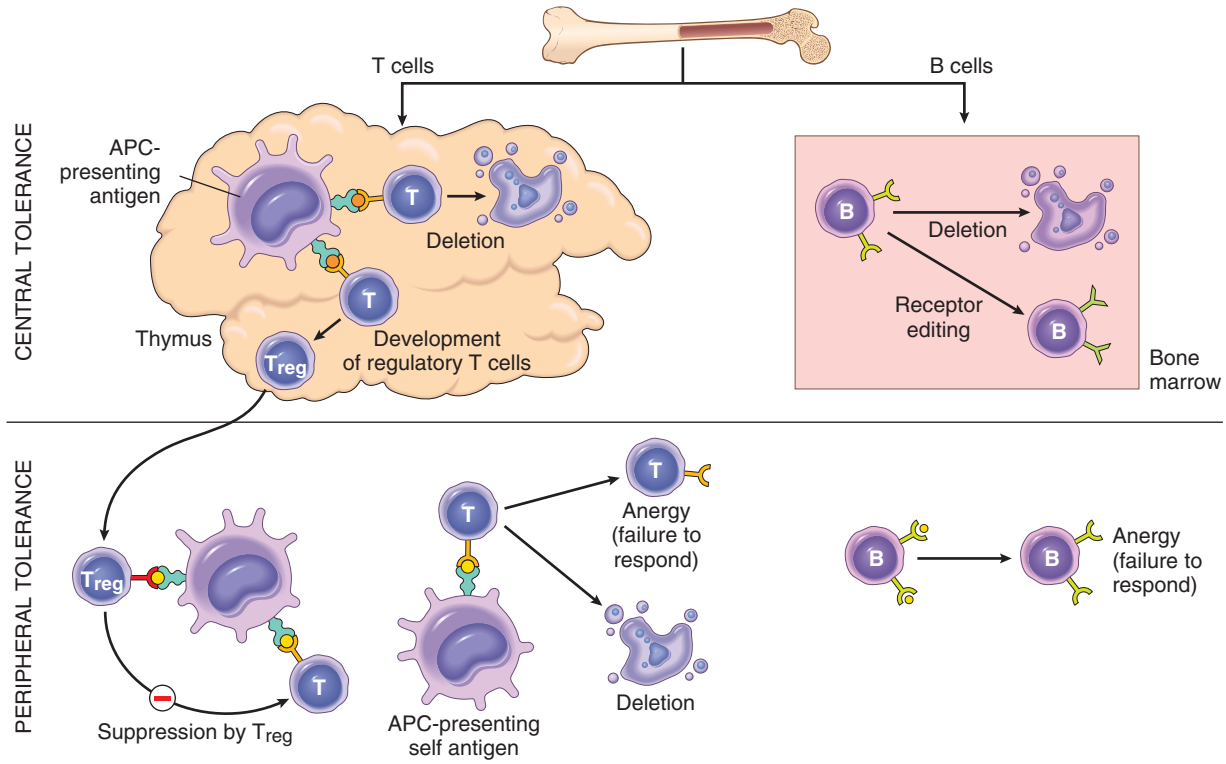


Figure 4–15 Immunologic self-tolerance: The principal mechanisms of central and peripheral self-tolerance in T and B cells.

signals (e.g., through B7 molecules) provided by the APCs. If the second costimulatory signals are not delivered, or if an inhibitory receptor on the T cell (rather than the costimulatory receptor) is engaged when the cell encounters self antigen, the T cell becomes anergic and cannot respond to the antigen (Fig. 4–15). Because costimulatory molecules are not strongly expressed on most normal tissues, the encounter between autoreactive T cells and self antigens in tissues may result in anergy. B cells can also become anergic if they encounter antigen in the absence of specific helper T cells.

- *Suppression by regulatory T cells*: The responses of T lymphocytes to self antigens may be actively suppressed by regulatory T cells. The best-defined populations of regulatory T cells express CD25, one of the chains of the receptor for IL-2, and require IL-2 for their generation and survival. These cells also express a unique transcription factor called FoxP3. This protein is necessary for the development of regulatory cells, and mutations in the *FOXP3* gene are responsible for a systemic autoimmune disease called IPEX (immune dysregulation, polyendocrinopathy, enteropathy, X-linked syndrome), which is associated with deficiency of regulatory T cells. Several mechanisms have been proposed to explain how regulatory T cells control immune responses, including secretion of immunosuppressive cytokines (e.g., IL-10, transforming growth factor-β [TGF-β]), which can dampen a variety of T cell responses, and competitive blocking of B7 molecules on APCs.
- *Activation-induced cell death*: Another mechanism of peripheral tolerance involves apoptosis of mature lymphocytes as a result of self-antigen recognition. One mechanism of apoptosis involves the death receptor Fas (a member of the TNF receptor family), which can be engaged by its ligand coexpressed on the same or neighboring cells. The same pathway is important for the deletion of self-reactive B cells by Fas ligand expressed on T cells. The importance of this pathway of self-tolerance is illustrated by the discovery that mutations in the *FAS* gene are responsible for an autoimmune disease called the autoimmune lymphoproliferative syndrome (ALPS), characterized by lymphadenopathy and multiple autoantibodies including anti-DNA. Defects in Fas and Fas ligand are also the cause of similar autoimmune diseases in mice. The mitochondrial pathway of apoptosis, which does not depend on death receptors, may also be involved in the elimination of self-reactive lymphocytes.

Mechanisms of Autoimmunity

Proceeding from the foregoing summary of the principal mechanisms of self-tolerance, we can ask how these mechanisms might break down to give rise to pathologic autoimmunity. Unfortunately, there are no simple answers to this question, and the underlying causes of most human autoimmune diseases remain to be determined. As mentioned earlier, certain mutations can compromise one or another pathway of self-tolerance and cause pathologic autoimmunity. Study of these single-gene mutations is extremely informative, and such research helps to establish the biologic significance of the various pathways of self-tolerance. The diseases caused by such mutations are rare, however, and most autoimmune diseases cannot be explained by defects in single genes.

It is believed that *the breakdown of self-tolerance and development of autoimmunity result from a combination of inherited susceptibility genes, which influence lymphocyte tolerance, and environmental factors, such as infections or tissue injury, that alter the display of self antigens* (Fig. 4–16).

Genetic Factors in Autoimmunity

There is abundant evidence that susceptibility genes play an important role in the development of autoimmune diseases.

- Autoimmune diseases have a tendency to run in families, and there is a greater incidence of the same disease in monozygotic than in dizygotic twins.
- Several autoimmune diseases are linked with the HLA locus, especially class II alleles (HLA-DR, -DQ). The frequency of a disease in a person with a particular HLA allele, compared with that in people who do not inherit that allele, is called the *odds ratio* or *relative risk* (Table 4–7). The relative risk ranges from 3 or 4 for rheumatoid arthritis (RA) and HLA-DR4 to 100 or more for ankylosing spondylitis and HLA-B27. However, how MHC genes influence the development of autoimmunity is still not clear, especially because MHC molecules do not distinguish between self and foreign peptide antigens. It is also worthy of note that most people with a susceptibility-related MHC allele never develop any disease, and, conversely, people without the relevant MHC gene can develop the disease. Expression of a particular MHC gene is therefore but one variable that can contribute to autoimmunity.
- Genome-wide association studies and linkage studies in families are revealing many genetic polymorphisms that are associated with different autoimmune diseases (Table 4–8). Some of these polymorphisms seem to be associated with several diseases, suggesting that the genes involved influence general mechanisms of self-tolerance and immune regulation. Others are disease-specific and may influence end-organ sensitivity or display of particular self antigens. There is great interest in elucidating how these genes contribute to autoimmunity, and many plausible hypotheses have been proposed (Table 4–8), but the actual role of these genes in the development of particular autoimmune diseases is not established.

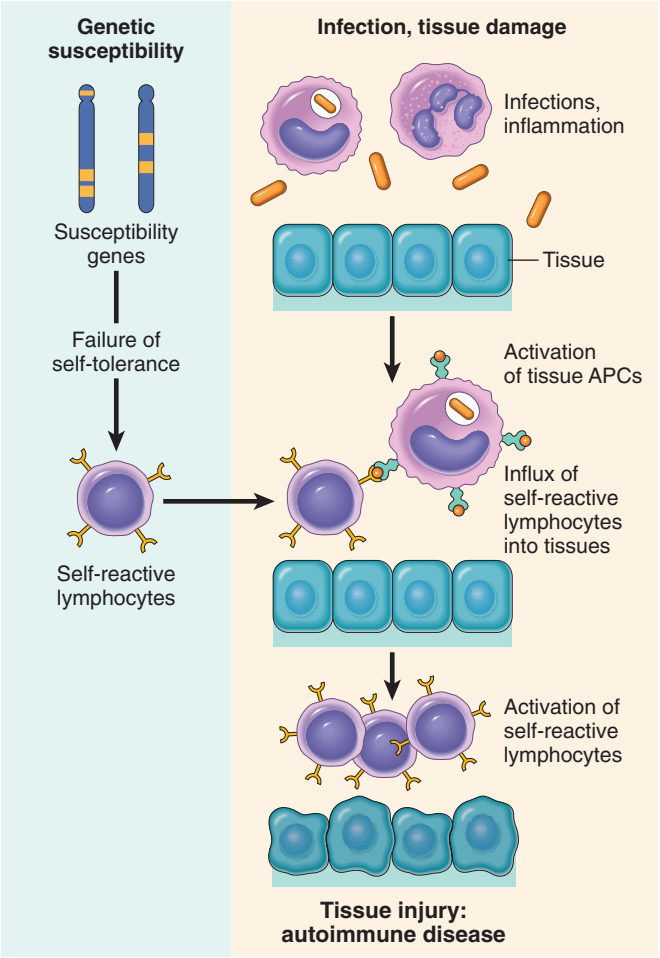


Figure 4–16 Pathogenesis of autoimmunity. Autoimmunity arises from the inheritance of susceptibility genes that may interfere with self-tolerance, in association with environmental triggers (infection, tissue injury, inflammation) that alter the display of self antigens, promote lymphocyte entry into tissues, and enhance the activation of self-reactive lymphocytes.

Role of Infections and Tissue Injury

A variety of microbes, including bacteria, mycoplasmas, and viruses, have been implicated as triggers for autoimmunity. Microbes may induce autoimmune reactions by several mechanisms:

Table 4–7 Association of Human Leukocyte Antigen (HLA) Alleles with Autoimmune Diseases

Disease	HLA Allele	Odds Ratio*
Rheumatoid arthritis (anti-CCP Ab–positive)†	DRB1	4–12
Type 1 diabetes	DRB1*0301-DQA1*0501-DQB1*0201 haplotype	4
	DRB1*0401- DQA1*0301-DQB1*0302 haplotype	8
	DRB1*0301/0401 haplotype heterozygotes	35
Multiple sclerosis	DRB1*1501	3
Systemic lupus erythematosus	DRB1*0301	2
	DRB1*1501	1.3
Ankylosing spondylitis	B*27 (mainly B*2705 and B*2702)	100–200
Celiac disease	DQA1*0501-DQB1*0201 haplotype	7

*The *odds ratio* (also called *relative risk*) is the approximate value of the increased risk of the disease associated with the inheritance of particular HLA alleles. The data are from European-derived populations.
†Anti-CCP Ab, antibodies directed against cyclic citrullinated peptides. Data are from patients who tested positive for these antibodies in serum.
Table courtesy of Dr. Michelle Fernando, Imperial College London.

Table 4–8 Selected Non–Human Leukocyte Antigen (HLA) Genes Associated with Autoimmune Diseases

Putative Gene Involved*	Diseases	Postulated Function of Encoded Protein and Role of Mutation/Polymorphism in Disease
Genes Involved in Immune Regulation		
PTPN22	RA, T1D, IBD	Protein tyrosine phosphatase, may affect signaling in lymphocytes and may alter negative selection or activation of self-reactive T cells
IL23R	IBD, PS, AS	Receptor for the T_H17-inducing cytokine IL-23; may alter differentiation of CD4+ T cells into pathogenic T_H17 effector cells
CTLA4	T1D, RA	Inhibits T cell responses by terminating activation and promoting activity of regulatory T cells; may interfere with self-tolerance
IL2RA	MS, T1D	α chain of the receptor for IL-2, which is a growth and survival factor for activated and regulatory T cells; may affect development of effector cells and/or regulation of immune responses
Genes Involved in Immune Responses to Microbes		
NOD2	IBD	Cytoplasmic sensor of bacteria expressed in Paneth and other intestinal epithelial cells; may control resistance to gut commensal bacteria
ATG16	IBD	Involved in autophagy; possible role in defense against microbes and maintenance of epithelial barrier function
IRF5, IFIH1	SLE	Role in production of type I IFN, involved in the pathogenesis of SLE (see text)

*The probable linkage of these genes with various autoimmune diseases has been defined by genome-wide association studies (GWAS) and other methods for studying disease-associated polymorphisms.
AS, ankylosing spondylitis; IBD, inflammatory bowel disease; IFN, interferon; MS, multiple sclerosis; PS, psoriasis; RA, rheumatoid arthritis; SLE, systemic lupus erythematosus; T1D, type 1 diabetes.
Adapted from Zenewicz L, Abraham C, Flavell RA, Cho J: Unraveling the genetics of autoimmunity. Cell 140:791, 2010.

- *Viruses and other microbes may share cross-reacting epitopes with self antigens,* such that responses may be induced by the microbe but may attack self tissues. This phenomenon is called molecular mimicry. It is the probable cause of a few diseases, the best example being rheumatic heart disease, in which an immune response against streptococci cross-reacts with cardiac antigens. It is not known if more subtle mimicry plays a role in other autoimmune diseases.
- *Microbial infections with resultant tissue necrosis and inflammation can cause upregulation of costimulatory molecules on APCs in the tissue,* thus favoring a breakdown of T cell anergy and subsequent T cell activation.

There is no lack of possible mechanisms to explain how infectious agents might participate in the pathogenesis of autoimmunity. At present, however, no evidence is available that clearly implicates any microbe in the causation of human autoimmune diseases. Adding to the complexity are recent suggestions (based largely on epidemiologic data) that infections may paradoxically protect affected persons from some autoimmune diseases, notably type 1 diabetes and multiple sclerosis. The possible mechanisms underlying this effect are not understood.

The display of tissue antigens may be altered by a variety of environmental insults, not only infections. As discussed later, ultraviolet (UV) radiation causes cell death and may lead to the exposure of nuclear antigens, which elicit pathologic immune responses in lupus; this mechanism is the proposed explanation for the association of lupus flares with exposure to sunlight. Smoking is a risk factor for RA, perhaps because it leads to chemical modification of self antigens. Local tissue injury for any reason may lead to the release of self antigens and autoimmune responses.

Finally, there is a strong gender bias of autoimmunity, with many of these diseases being more common in women than in men. The underlying mechanisms are still not well understood, and may include the effects of hormones and other factors.

An autoimmune response may itself promote further autoimmune attack. Tissue injury caused by an autoimmune response or any other cause may lead to exposure of self antigen epitopes that were previously concealed but are now presented to T cells in an immunogenic form. The activation of such autoreactive T cells is called "epitope spreading," because the immune response "spreads" to epitopes that were not recognized initially. This is one of the mechanisms that may contribute to the chronicity of autoimmune diseases.

SUMMARY

Immunologic Tolerance and Autoimmunity

- *Tolerance* (unresponsiveness) to self antigens is a fundamental property of the immune system, and breakdown of tolerance is the basis of autoimmune diseases.
- *Central tolerance*: Immature lymphocytes that recognize self antigens in the central (generative) lymphoid organs are killed by apoptosis; in the B cell lineage, some of the self-reactive lymphocytes switch to new antigen receptors that are not self-reactive.
- *Peripheral tolerance*: Mature lymphocytes that recognize self antigens in peripheral tissues become functionally inactive (anergic), or are suppressed by regulatory T lymphocytes, or die by apoptosis.
- The *factors that lead to a failure of self-tolerance and the development of autoimmunity* include (1) inheritance of susceptibility genes that may disrupt different tolerance pathways and (2) infections and tissue alterations that may expose self-antigens and activate APCs and lymphocytes in the tissues.

Having discussed the general principles of tolerance and autoimmunity, we proceed to a discussion of some of the most common and important autoimmune diseases. Although each disease is discussed separately, considerable overlap is apparent in their clinical, serologic, and morphologic features. Only the systemic autoimmune diseases are covered in this chapter; the autoimmune diseases that affect single organ systems are more appropriately discussed in the chapters that deal with the relevant organs.

Systemic Lupus Erythematosus

Systemic lupus erythematosus (SLE) is a multisystem autoimmune disease of protean manifestations and variable clinical behavior. Clinically, it is an unpredictable, remitting and relapsing disease of acute or insidious onset that may involve virtually any organ in the body; however, it affects principally the skin, kidneys, serosal membranes, joints, and heart. Immunologically, the disease is associated with an enormous array of autoantibodies, classically including antinuclear antibodies (ANAs). The clinical presentation of SLE is so variable, with so many overlapping features with those of other autoimmune diseases (RA, polymyositis, and others), that it has been necessary to develop diagnostic criteria for SLE (Table 4–9). The diagnosis is established by demonstration of four or more of the criteria during any interval of observation.

Incidence and prevalence estimates of SLE vary among racial and ethnic groups; some studies estimate the prevalence to be as high as 0.2% in certain groups. As with many autoimmune diseases, there is a strong (approximately 9:1) female preponderance, and the disease affects 1 in 700 women of childbearing age. SLE is more common and severe in black Americans, affecting 1 in 245 women in that group. Onset typically is in the second or third decade of life, but it may manifest at any age, including early childhood.

PATHOGENESIS

The fundamental defect in SLE is a failure to maintain self-tolerance, leading to the production of a large number of autoantibodies that can damage tissues either directly or in the form of immune complex deposits. As in other autoimmune diseases, the pathogenesis of SLE involves a combination of genetic and environmental factors. Recent studies have revealed interesting clues about the pathogenesis of this enigmatic disorder (Fig. 4–17).

Genetic Factors. Many lines of evidence support a genetic predisposition to SLE.

- **Familial association.** Family members have an increased risk for the development of SLE, and up to 20% of clinically unaffected first-degree relatives may have autoantibodies. There is a high rate of concordance in

Table 4–9 1997 Revised Criteria for Classification of Systemic Lupus Erythematosus*

Criterion	Definition
1. Malar rash	Fixed erythema, flat or raised, over the malar eminences, tending to spare the nasolabial folds
2. Discoid rash	Erythematous raised patches with adherent keratotic scaling and follicular plugging; atrophic scarring may occur in older lesions
3. Photosensitivity	Rash occurring as an unusual reaction to sunlight, reported in patient history or as physician observation
4. Oral ulcers	Oral or nasopharyngeal ulceration, usually painless, observed by a physician
5. Arthritis	Nonerosive arthritis involving two or more peripheral joints, characterized by tenderness, swelling, or effusion
6. Serositis	Pleuritis—convincing history of pleuritic pain or rub heard by a physician or evidence of pleural effusion *or* Pericarditis—documented by electrocardiogram or rub or evidence of pericardial effusion
7. Renal disorder	Persistent proteinuria >0.5 g/dL or >3+ if quantitation not performed *or* Cellular casts—may be red blood cell, hemoglobin, granular, tubular, or mixed
8. Neurologic disorder	Seizures—in the absence of offending drugs or known metabolic derangements, (e.g., uremia, ketoacidosis, or electrolyte imbalance) *or* Psychosis—in the absence of offending drugs or known metabolic derangements, (e.g., uremia, ketoacidosis, or electrolyte imbalance)
9. Hematologic disorder	Hemolytic anemia—with reticulocytosis *or* Leukopenia—$<4.0 \times 10^9$/L ($4000/mm^3$) total on two or more occasions *or* Lymphopenia—$<1.5 \times 10^9$/L ($1500/mm^3$) on two or more occasions *or* Thrombocytopenia—$<100 \times 10^9$/L ($100 \times 10^3/mm^3$) in the absence of offending drugs
10. Immunologic disorder	Anti-DNA antibody to native DNA in abnormal titer or Anti-Sm—presence of antibody to Sm nuclear antigen or Positive finding of antiphospholipid antibodies based on (1) an abnormal serum level of IgG or IgM anticardiolipin antibodies, (2) a positive test for lupus anticoagulant using a standard test, or (3) a false-positive serologic test for syphilis known to be positive for at least 6 months and confirmed by negative *Treponema pallidum* immobilization or fluorescent treponemal antibody absorption test
11. Antinuclear antibody	An abnormal titer of antinuclear antibody by immunofluorescence or an equivalent assay at any point in time and in the absence of drugs known to be associated with drug-induced lupus syndrome

*The proposed classification is based on 11 criteria. For the purpose of identifying patients in clinical studies, a person is said to have systemic lupus erythematosus if any 4 or more of the 11 criteria are present, serially or simultaneously, during any period of observation.

From Tan EM, Cohen AS, Fries JF, et al: The revised criteria for the classification of systemic lupus erythematosus. Arthritis Rheum 25:1271, 1982; and Hochberg MC: Updating the American College of Rheumatology revised criteria for the classification of systemic lupus erythematosus. Arthritis Rheum 40:1725, 1997.

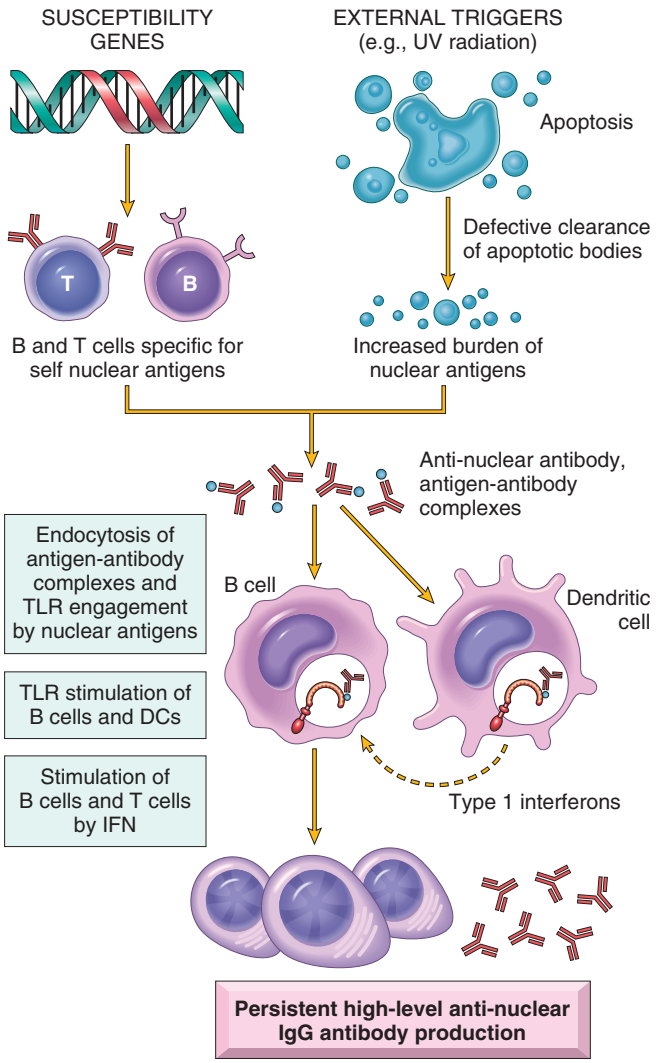


Figure 4–17 Model for the pathogenesis of systemic lupus erythematosus. Genetic susceptibility and exposure result in failure of self-tolerance and persistence of nuclear antigens. Autoantibodies serve to internalize nuclear components, which engage TLRs and stimulate IFN production. IFN may stimulate B and T cell responses to the nuclear antigens. IFN, interferon; IgG, immunoglobulin G; MHC, major histocompatibility complex; TLRs, Toll-like receptors; UV, ultraviolet.

monozygotic twins (25%) versus dizygotic twins (1% to 3%).

- **HLA association.** The odds ratio (relative risk) for persons with HLA-DR2 or HLA-DR3 is 2 to 3, and if both haplotypes are present, the risk is about 5.
- **Other genes.** Genetic deficiencies of classical pathway complement proteins, especially C1q, C2, or C4, are seen in about 10% of patients with SLE. The complement deficiencies may result in defective clearance of immune complexes and apoptotic cells, and failure of B cell tolerance. A polymorphism in the inhibitory Fc receptor, FcγRIIb, has been described in some patients; this may contribute to inadequate control of B cell activation. Many other genes have been detected by genome-wide association studies, but the role of each of these is not established and their contribution to the development of the disease remains unclear.

Environmental Factors. There are many indications that environmental factors are involved in the pathogenesis of SLE.

- **Ultraviolet (UV) radiation** (sun exposure) exacerbates the lesions of SLE. A postulated mechanism of this effect is that UV radiation causes apoptosis of host cells, leading to an increased burden of nuclear fragments and inflammatory responses to the products of dead cells.
- **Cigarette smoking** has been shown to be associated with the development of SLE. Although the mechanism of this is unknown, smoking tobacco may modulate the production of autoantibodies.
- **Sex hormones** had been thought to exert an important influence on the development of disease, because SLE is 10 times more common in women during reproductive years than in men of similar ages but only 2 to 3 times more common in women during childhood or after the age of 65. However, treatment of women with oral contraceptives containing high doses of estrogen and progesterone did not influence the frequency or severity of disease flares, suggesting that factors other than hormones may account for the increased risk of this disease in women.
- **Drugs** such as procainamide and hydralazine can induce an SLE-like disease, although typically glomerulonephritis does not develop. These drugs cause demethylation of DNA, which could influence the expression of a variety of genes involved in the development of autoimmunity, or the ability of DNA to activate host cells.

Immunologic Abnormalities in SLE. Studies have implicated several components of the innate and adaptive immune system in the pathogenesis of SLE.

- **Type I interferons.** Blood cells show a striking molecular signature that indicates exposure to interferon-α (IFN-α), a type I interferon that is produced mainly by plasmacytoid DCs. Some studies have shown that such cells from patients with SLE also produce abnormally large amounts of IFN-α.
- **TLR signals.** Studies in animal models have shown that TLRs that recognize DNA and RNA, notably the DNA-recognizing TLR9 and the RNA-recognizing TLR7, produce signals that activate B cells specific for self nuclear antigens.
- **Failure of B cell tolerance.** Studies with B cells from patients with SLE suggest the presence of defects in both central and peripheral tolerance, resulting in a higher frequency of autoreactive B cells than that typical for healthy people.

Based on these clues, a model for the pathogenesis of SLE has been proposed (Fig. 4–17). According to this model, UV irradiation and other environmental insults lead to the apoptosis of cells. Inadequate clearance of the nuclei of these cells, in part because of defects in clearance mechanisms such as complement proteins and receptors, results in a large burden of nuclear antigens. Polymorphisms in various genes, which are susceptibility genes for lupus, lead to defective ability to maintain self-tolerance in B and T lymphocytes, because of which self-reactive lymphocytes remain functional. The self-reactive B cells are stimulated by the self nuclear antigens, and antibodies are produced against the antigens. Complexes of the antigens and antibodies bind to

Fc receptors on B cells and DCs and may be internalized. The nucleic acid components engage TLRs and stimulate B cells to produce autoantibodies and activate DCs, particularly plasmacytoid DCs, to produce IFN-α, which further enhances the immune response and causes more apoptosis. The net result is a cycle of antigen release and immune activation resulting in the production of high-affinity autoantibodies.

Spectrum of Autoantibodies in SLE

Antibodies have been identified against a host of nuclear and cytoplasmic components of the cell that are specific to neither organs nor species. Another group of antibodies is directed against surface antigens of blood cells, while yet another is reactive with proteins in complex with phospholipids (antiphospholipid antibodies) (Chapter 3).

- *Antinuclear antibodies*. ANAs are directed against several nuclear antigens and can be grouped into four categories: (1) antibodies to DNA, (2) antibodies to histones, (3) antibodies to nonhistone proteins bound to RNA, and (4) antibodies to nucleolar antigens. Table 4–10 lists several autoantibodies, including ANAs, and their association with SLE as well as with other autoimmune diseases, to be discussed later. The most widely used method of detecting ANAs is the indirect immunofluorescence assay (IFA), which screens for autoantibodies that bind to a variety of nuclear antigens, including DNA, RNA, and proteins. Four staining patterns are seen with IFA: homogeneous or diffuse, rim or peripheral, speckled, and nucleolar. While each pattern can be suggestive of the presence of specific autoantibodies, the strength of these associations is limited and should not be relied on. *ANA testing by IFA is extremely sensitive, as more than 95% of patients with SLE will test positive, but the test's specificity is quite limited, because patients with other autoimmune diseases, chronic infections, and cancer will test positive as well*. Furthermore, ANAs are seen in approximately 5% to 15% of healthy people, and the incidence increases with age. Recently, the IFA has been replaced in many clinical laboratories by multiplex flow cytometry immunoassays that can simultaneously test for multiple specific autoantibodies, but these assays may lack the sensitivity of the IFA. *Antibodies to double-stranded DNA (dsDNA) and the so-called Smith (Sm) antigen can be detected by ELISA or multiplex flow methods and are specific for SLE.*
- *Other autoantibodies*. Antibodies against blood cells, including red cells, platelets, and lymphocytes, are found in many patients. Antiphospholipid antibodies are present in 40% to 50% of patients with lupus and react with a wide variety of proteins in complex with phospholipids. Some bind to cardiolipin antigen, used in serologic tests for syphilis, so patients with lupus may have a false-positive test result for syphilis. Antiphospholipid antibodies contribute to coagulation abnormalities, which are described below.

Mechanisms of Tissue Injury

Regardless of the exact sequence by which autoantibodies are formed, they are likely to be the mediators of tissue injury, probably through multiple mechanisms.

- *Most organ damage in SLE is caused by immune complex deposition.* Skin and kidney biopsies from patients with SLE typically demonstrate diffuse and heavy granular deposits of complement and immunoglobulin. Autoantibodies complexed with DNA can be detected as well. These deposits of immune complexes had been thought to cause tissue damage by activating the classical complement pathway (type III hypersensitivity); 75% of patients will have reduced serum levels of C3 and C4 at the time of SLE flares, presumably because complement is being activated and consumed faster than it can be produced. However, people and rodents deficient in C1q are not protected from SLE and actually can spontaneously develop SLE, raising the possibility that

Table 4–10 Selected Autoantibodies Associated with Presumed Autoimmune Diseases

Autoantibody (Specificity)	Major Disease Association(s)	Likely Role(s) in Disease
Anti-dsDNA (double-stranded DNA)	SLE*	Formation of immune complexes
Anti-Sm (ribonuclear core protein, Sm antigen)	SLE*	Formation of immune complexes
Anti-RNP U1 (ribonuclear protein)	SLE, mixed connective tissue disease	Formation of immune complexes
Anti–SS-A (Ro), anti–SS-B (La) (ribonucleoproteins)	Sjögren syndrome, SLE	Role in Sjögren syndrome not known
Anti–Scl-70 (DNA topoisomerase I)	Systemic sclerosis*	Unknown
Anti-histones (histone proteins)	SLE	Formation of immune complexes
Anti-centromere (centromere proteins)	Limited scleroderma, systemic sclerosis*	Unknown
Antiphospholipid (phospholipid–protein complexes involved in blood coagulation)	Antiphospholipid syndrome, SLE	Thrombotic episodes
Anti-Jo1 (histidyl tRNA ligase)	Inflammatory myopathies*	Unknown
Anti-mitochondrial	Primary biliary cirrhosis*	Unknown
Anti-eTg (transglutaminase)	Dermatitis herpetiformis	Unknown
Anti–neutrophil cytoplasmic antibody (ANCA) (proteins in neutrophil cytoplasm)	Various vasculitides*	Formation of immune complexes?
		Neutrophil degranulation?
Anti–smooth muscle	Chronic autoimmune hepatitis	Unknown

Each antibody specificity is detected in 30% to 90% of patients with a particular disease. Asterisks indicate high correlation between the antibody specificity and the disease.
SLE, systemic lupus erythematosus.

complement-independent mechanisms may also contribute to tissue damage.

- *Autoantibodies of different specificities contribute to the pathology and clinical manifestations of SLE* (type II hypersensitivity). *Autoantibodies against red cells, white cells, and platelets* opsonize these cells and lead to their phagocytosis, resulting in cytopenias. *Autoantibodies against various phospholipids* lead to increased thrombosis in patients, with varied clinical consequences, including recurrent spontaneous abortion and thrombotic episodes. These disorders are part of the *antiphospholipid syndrome*. Paradoxically, these antibodies interfere with clotting tests and are actually called "lupus anticoagulants." Autoantibodies are also produced against clotting factors such as thrombin, and these too may contribute to clotting disorders. *Autoantibodies against central nervous system receptors for various neurotransmitters* have been implicated in the neuropsychiatric complications of the disease.
- There is no evidence that ANAs can permeate intact cells. However, if cell nuclei are exposed, the ANAs can bind to them. In tissues, nuclei of damaged cells react with ANAs, lose their chromatin pattern, and become homogeneous, to produce so-called *LE bodies* or *hematoxylin bodies*. An in vitro correlate of this is the *LE cell,* a neutrophil or macrophage that has engulfed the denatured nucleus of another injured cell. When blood is withdrawn and agitated, a number of leukocytes are sufficiently damaged to expose their nuclei to ANAs, with secondary complement activation; these antibody- and complement-opsonized nuclei are then readily phagocytosed. Although the LE cell test is positive in as many as 70% of patients with SLE, it is now largely of historical interest.

MORPHOLOGY

SLE is a systemic disease with protean manifestations (Table 4–9). The morphologic changes in SLE are therefore extremely variable and depend on the nature of the autoantibodies, the tissue in which immune complexes deposit, and the course and duration of disease. The most characteristic morphologic changes result from the deposition of immune complexes in a variety of tissues.

Blood Vessels. An **acute necrotizing vasculitis** affecting small arteries and arterioles may be present in any tissue. The arteritis is characterized by necrosis and by fibrinoid deposits within vessel walls containing antibody, DNA, complement fragments, and fibrinogen; a transmural and perivascular leukocytic infiltrate is also frequently present. In chronic stages, vessels show fibrous thickening with luminal narrowing.

Kidneys. Kidney involvement is one of the most important clinical features of SLE, with renal failure being the most common cause of death. The focus here is on glomerular pathology, although interstitial and tubular lesions are also seen in SLE.

The pathogenesis of all forms of **glomerulonephritis** in SLE involves deposition of DNA–anti-DNA complexes within the glomeruli. These evoke an inflammatory response that may cause proliferation of the endothelial, mesangial, and/or epithelial cells and, in severe cases, necrosis of the glomeruli. Although the kidney appears normal by light microscopy in 25% to 30% of cases, almost all cases of SLE show some renal abnormality if examined by immunofluorescence and electron microscopy. According to the current International Society of Nephrology/Renal Pathology Society morphologic classification, there are six patterns of glomerular disease in SLE (none of which is specific to the disease): **class I,** minimal mesangial lupus nephritis; **class II,** mesangial proliferative lupus nephritis; **class III,** focal lupus nephritis; **class IV,** diffuse lupus nephritis; **class V,** membranous lupus nephritis; and **class VI,** advanced sclerosing lupus nephritis.

- **Minimal mesangial lupus nephritis (class I)** is rarely encountered in renal biopsies. Immune complexes are present in the mesangium, but there are no concomitant structural alterations detectable by light microscopy.
- **Mesangial proliferative lupus nephritis (class II)** is seen in 10% to 25% of cases and is associated with mild clinical symptoms. Immune complexes deposit in the mesangium, with a mild to moderate increase in the mesangial matrix and cellularity.
- **Focal lupus nephritis (class III)** is seen in 20% to 35% of cases. Lesions are visualized in fewer than half the glomeruli, and they may be segmentally or globally distributed within each glomerulus. Active lesions are characterized by swelling and proliferation of endothelial and mesangial cells, infiltration by neutrophils, and/or fibrinoid deposits with capillary thrombi (Fig. 4–18, *A*). The clinical presentation may range from only mild microscopic hematuria and proteinuria to a more active urinary sediment with red blood cell casts and acute, severe renal insufficiency.
- **Diffuse lupus nephritis (class IV)** is the most serious form of renal lesions in SLE and is also the most commonly encountered in renal biopsies, occurring in 35% to 60% of patients. It is distinguished from focal lupus nephritis (class III) by involvement of half or more of glomeruli. Most of the glomeruli show endothelial and mesangial proliferation, leading to diffuse hypercellularity of these structures (Fig. 4–18, *B*) and producing in some cases epithelial crescents that fill Bowman's space. When extensive, subendothelial immune complexes create a circumferential thickening of the capillary wall, resembling rigid "wire loops" on routine light microscopy (Fig. 4–18, *C*). Electron microscopy reveals prominent electron-dense subendothelial immune complexes (between endothelium and basement membrane) (Fig. 4–18, *D*), but immune complexes are also usually present in other parts of the capillary wall and in the mesangium. Immune complexes can be visualized by staining with fluorescent antibodies directed against immunoglobulins or complement, resulting in a granular fluorescent staining pattern (Fig. 4–18, *E*). In due course, glomerular injury may give rise to scarring (glomerulosclerosis). Most affected patients have hematuria with moderate to severe proteinuria, hypertension, and renal insufficiency.
- **Membranous lupus nephritis (class V)** occurs in 10% to 15% of cases and is the designation for glomerular disease characterized by widespread thickening of the capillary wall due to deposition of subepithelial immune complexes. Membranous glomerulonephritis associated

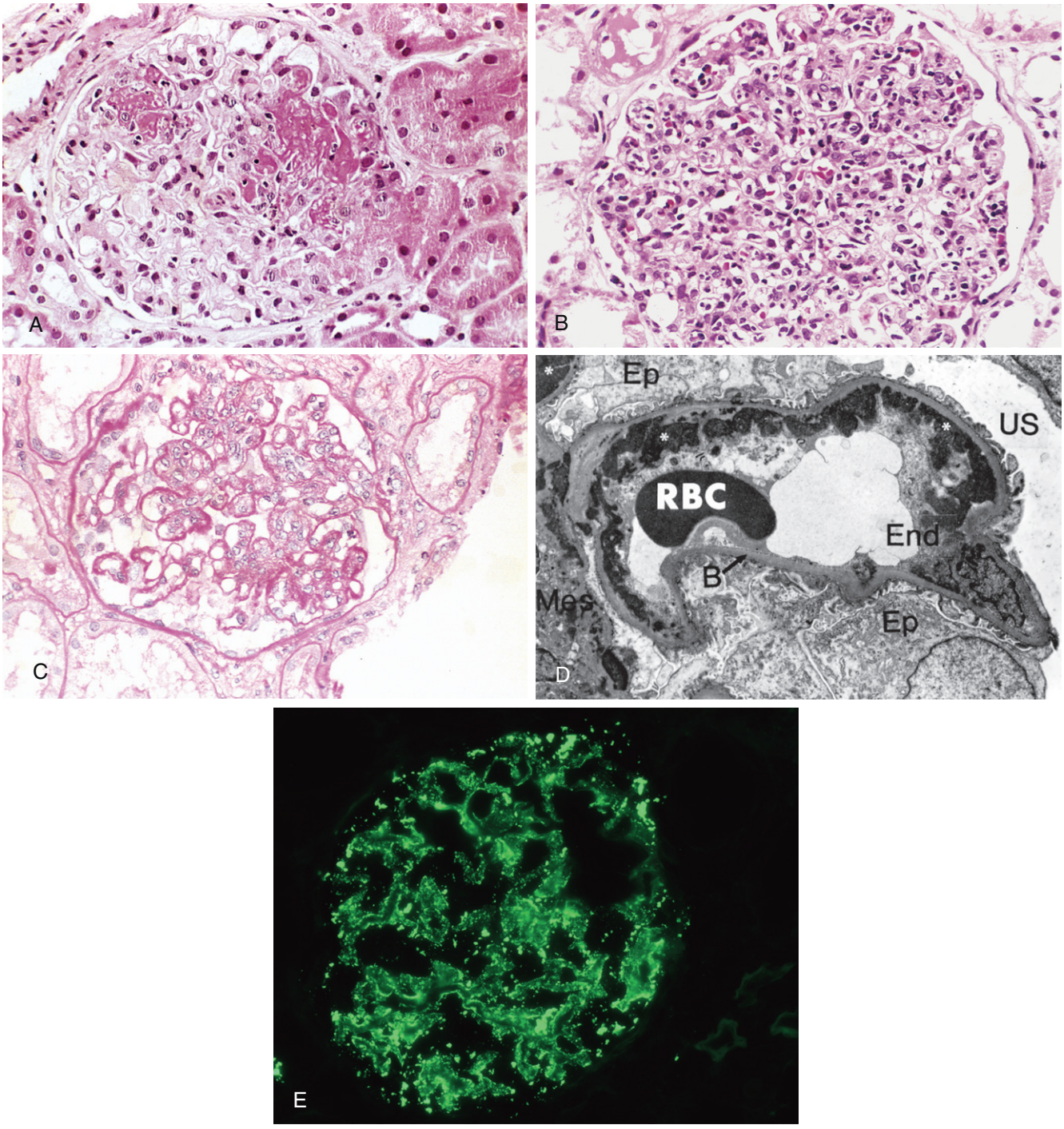


Figure 4–18 Lupus nephritis. **A,** Focal lupus nephritis, with two necrotizing lesions in a glomerulus (segmental distribution) (H&E stain). **B,** Diffuse lupus nephritis. Note the marked global increase in cellularity throughout the glomerulus (H&E stain). **C,** Lupus nephritis showing a glomerulus with several "wire loop" lesions representing extensive subendothelial deposits of immune complexes (periodic acid Schiff stain). **D,** Electron micrograph of a renal glomerular capillary loop from a patient with SLE nephritis. Confluent subendothelial dense deposits correspond to "wire loops" seen by light microscopy. **E,** Deposition of IgG antibody in a granular pattern, detected by immunofluorescence. B, basement membrane; End, endothelium; Ep, epithelial cell with foot processes; Mes, mesangium; RBC, red blood cell in capillary lumen; US, urinary space; *, electron-dense deposits in subendothelial location.

(A–C, courtesy of Dr. Helmut Rennke, Department of Pathology, Brigham and Women's Hospital, Boston, Massachusetts. D, Courtesy of Dr. Edwin Eigenbrodt, Department of Pathology, University of Texas Southwestern Medical School, Dallas. E, Courtesy of Dr. Jean Olson, Department of Pathology, University of California, San Francisco, California.)

with SLE is very similar to that encountered in idiopathic membranous nephropathy (Chapter 13). Thickening of capillary walls is caused by increased deposition of basement membrane–like material, as well as accumulation of immune complexes. Patients with this histologic change almost always have severe proteinuria with overt nephrotic syndrome (Chapter 13).

- **Advanced sclerosing lupus nephritis (class VI)** is characterized by **complete sclerosis** of greater than 90% of glomeruli and corresponds to clinical end stage renal disease.

Skin. The **skin** is involved in a majority of patients; a characteristic erythematous or maculopapular eruption over the malar eminences and bridge of the nose ("butterfly pattern") is observed in approximately half of the cases. Exposure to sunlight (UV light) exacerbates the erythema (so-called **photosensitivity**), and a similar rash may be present elsewhere on the extremities and trunk, frequently in sun-exposed areas. Histopathologic findings include liquefactive degeneration of the basal layer of the epidermis, edema at the dermoepidermal junction, and mononuclear infiltrates around blood vessels and skin appendages (Fig. 4–19, *A*). Immunofluorescence microscopy reveals deposition of immunoglobulin and complement at the dermoepidermal junction (Fig. 4–19, *B*); similar immunoglobulin and complement deposits may also be present in apparently uninvolved skin.

Joints. Joint involvement is frequent but usually is not associated with striking anatomic changes or with joint deformity. When present, it consists of swelling and a nonspecific mononuclear cell infiltration in the synovial membranes. Erosion of the membranes and destruction of articular cartilage, such as in RA, are exceedingly rare.

CNS. Central nervous system (CNS) involvement also is very common, with focal neurologic deficits and/or neuropsychiatric symptoms. CNS disease often is ascribed to vascular lesions causing ischemia or multifocal cerebral microinfarcts. Small vessel angiopathy with noninflammatory intimal proliferation is the most frequent pathological lesion; frank vasculitis is uncommon. The angiopathy may result from thrombosis caused by antiphospholipid antibodies. Premature atherosclerosis occurs and may contribute to CNS ischemia. Another postulated mechanism for CNS disease is injury from antineuronal antibodies with consequent neurologic dysfunction, but this hypothesis remains unproved.

Other Organs. The **spleen** may be moderately enlarged. Capsular fibrous thickening is common, as is follicular hyperplasia with numerous plasma cells in the red pulp. Central penicilliary arteries characteristically show thickening and perivascular fibrosis, producing **onion-skin lesions**.

Pericardium and pleura, in particular, are **serosal membranes** that show a variety of inflammatory changes in SLE ranging (in the acute phase) from serous effusions to fibrinous exudates that may progress to fibrous opacification in the chronic stage.

Involvement of the heart is manifested primarily in the form of pericarditis. Myocarditis, in the form of a nonspecific mononuclear cell infiltrate, and valvular lesions, called

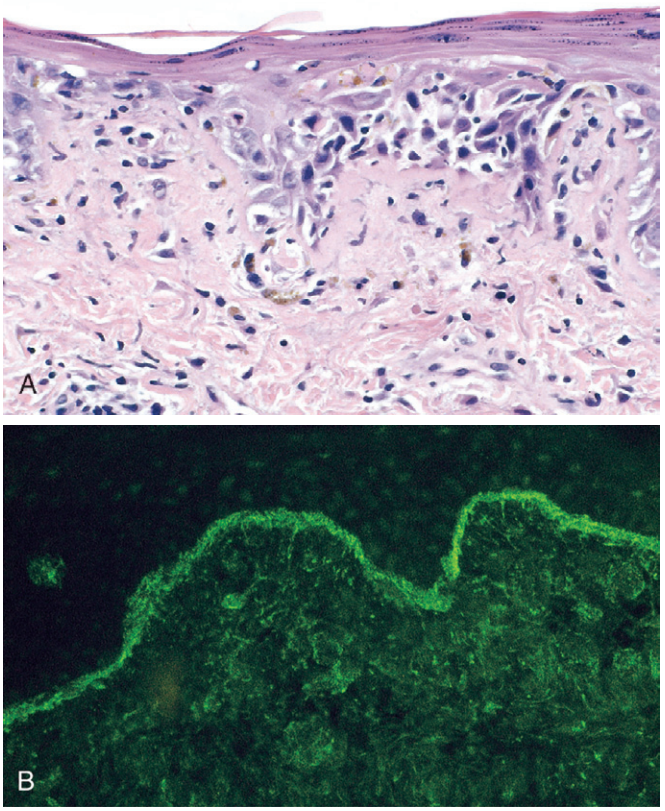


Figure 4–19 Systemic lupus erythematosus involving the skin. **A,** An H&E-stained section shows liquefactive degeneration of the basal layer of the epidermis and edema at the dermoepidermal junction. **B,** An immunofluorescence micrograph stained for IgG reveals deposits of immunoglobulin along the dermoepidermal junction. H&E, hematoxylin–eosin; IgG, immunoglobulin G.

(A, Courtesy of Dr. Jag Bhawan, Boston University School of Medicine, Boston, Massachusetts. B, Courtesy of Dr. Richard Sontheimer, Department of Dermatology, University of Texas Southwestern Medical School, Dallas, Texas.)

Libman-Sacks endocarditis, also occur but are less common in the current era of aggressive corticosteroid therapy. This **nonbacterial verrucous endocarditis** takes the form of irregular, 1- to 3-mm warty deposits, seen as distinctive lesions on either surface of the leaflets (i.e., on the surface exposed to the forward flow of the blood or on the underside of the leaflet) (see Chapter 10). An increasing number of patients also show clinical and anatomic manifestations of coronary artery disease. The basis of accelerated atherosclerosis is not fully understood, but the process seems to be multifactorial; certainly, immune complexes can deposit in the coronary vasculature, leading to endothelial damage by that pathway. Moreover, glucocorticoid treatment causes alterations in lipid metabolism, and renal disease (common in SLE) causes hypertension; both of these are risk factors for atherosclerosis (Chapter 9).

Many **other organs and tissues** may be involved. The changes consist essentially of acute vasculitis of the small vessels, foci of mononuclear infiltrations, and fibrinoid deposits. In addition, **lungs** may reveal interstitial fibrosis, along with pleural inflammation; the **liver** shows nonspecific inflammation of the portal tracts.

Clinical Manifestations

SLE is a multisystem disease that is highly variable in clinical presentation. Typically, the patient is a young woman with some, but rarely all, of the following features: a butterfly rash over the face, fever, pain and swelling in one or more peripheral joints (hands and wrists, knees, feet, ankles, elbows, shoulders), pleuritic chest pain, and photosensitivity. In many patients, however, the presentation of SLE is subtle and puzzling, taking forms such as a febrile illness of unknown origin, abnormal urinary findings, or joint disease masquerading as RA or rheumatic fever. ANAs are found in virtually 100% of patients, but an important point is that ANAs are not specific (Table 4–10). A variety of clinical findings may point toward renal involvement, including hematuria, red cell casts, proteinuria, and in some cases the classic nephrotic syndrome (Chapter 13). Laboratory evidence of some hematologic derangement is common, and in some patients anemia or thrombocytopenia may be the presenting manifestation as well as the dominant clinical problem. In still others, neuropsychiatric manifestations, including psychosis or convulsions, or coronary artery disease may be prominent clinical problems. Patients with SLE are also prone to infections, presumably because of their underlying immune dysfunction and treatment with immunosuppressive drugs. Recent strategies include B cell depletion with anti-CD20 antibody (Rituximab) and by blocking growth factors. The course of the disease is variable and unpredictable. Rare acute cases progress to death within weeks to months. More often, with appropriate therapy, the disease is characterized by flareups and remissions spanning a period of years or even decades. During acute flareups, increased deposition of immune complexes and the accompanying complement activation are thought to result in hypocomplementemia. Disease exacerbations usually are treated with corticosteroids or other immunosuppressive drugs. Even without therapy, in some patients the disease may run a benign course with only skin manifestations and mild hematuria for years. The outcome has improved significantly, and a 5-year survival can be expected in approximately 95% of patients. *The most common causes of death are renal failure, intercurrent infections, and cardiovascular disease.* The incidence of cancer also is increased, particularly B cell lymphomas, an association common to diseases marked by B cell hyperstimulation (e.g., Sjögren syndrome, discussed below). Patients treated with steroids and immunosuppressive drugs incur the usual risks associated with such therapy.

SUMMARY

Systemic Lupus Erythematosus

- SLE is a systemic autoimmune disease caused by autoantibodies produced against numerous self-antigens and the formation of immune complexes.
- The major autoantibodies, and the ones responsible for the formation of circulating immune complexes, are directed against nuclear antigens. Other autoantibodies react with erythrocytes, platelets, and various complexes of phospholipids with proteins.
- Disease manifestations include nephritis, skin lesions and arthritis (caused by the deposition of immune complexes), and hematologic and neurologic abnormalities.
- The underlying cause of the breakdown in self-tolerance in SLE is unknown; it may include excess or persistence of nuclear antigens, multiple inherited susceptibility genes, and environmental triggers (e.g., UV irradiation, which results in cellular apoptosis and release of nuclear proteins).

Rheumatoid Arthritis

Rheumatoid arthritis (RA) is a systemic, chronic inflammatory disease affecting many tissues but principally attacking the joints to produce a nonsuppurative proliferative synovitis that frequently progresses to destroy articular cartilage and underlying bone with resulting disabling arthritis. Because the principal lesions affect the joints and bones, this disease, as well as the juvenile form and other inflammatory diseases of joints, is discussed in Chapter 20.

Sjögren Syndrome

Sjögren syndrome is a clinicopathologic entity characterized by dry eyes (*keratoconjunctivitis sicca*) and dry mouth (*xerostomia*), resulting from immune-mediated destruction of the lacrimal and salivary glands. It occurs as an isolated disorder (primary form), also known as the *sicca syndrome,* or more often in association with another autoimmune disease (secondary form). Among the associated disorders, RA is the most common, but some patients have SLE, polymyositis, systemic sclerosis, vasculitis, or thyroiditis.

PATHOGENESIS

Several lines of evidence suggest that Sjögren syndrome is an autoimmune disease caused by CD4+ T cell reactions against unknown antigens in the ductal epithelial cells of the exocrine glands. There is also systemic B cell hyperactivity, as evidenced by the presence of ANAs and rheumatoid factor (RF) (even in the absence of associated RA). Most patients with primary Sjögren syndrome have autoantibodies to the ribonucleoprotein (RNP) antigens SS-A (Ro) and SS-B (La); note that these antibodies are also present in some SLE patients and are therefore not diagnostic for Sjögren syndrome (Table 4–10). Although patients with high-titer anti-SS-A antibodies are more likely to have systemic (extraglandular) manifestations, there is no evidence that the autoantibodies cause primary tissue injury. A viral trigger also has been suggested, but no causative virus has been identified conclusively. Genetic variables play a role in the pathogenesis of Sjögren syndrome. As with SLE, inheritance of certain class II MHC alleles predisposes to the development of specific RNP autoantibodies.

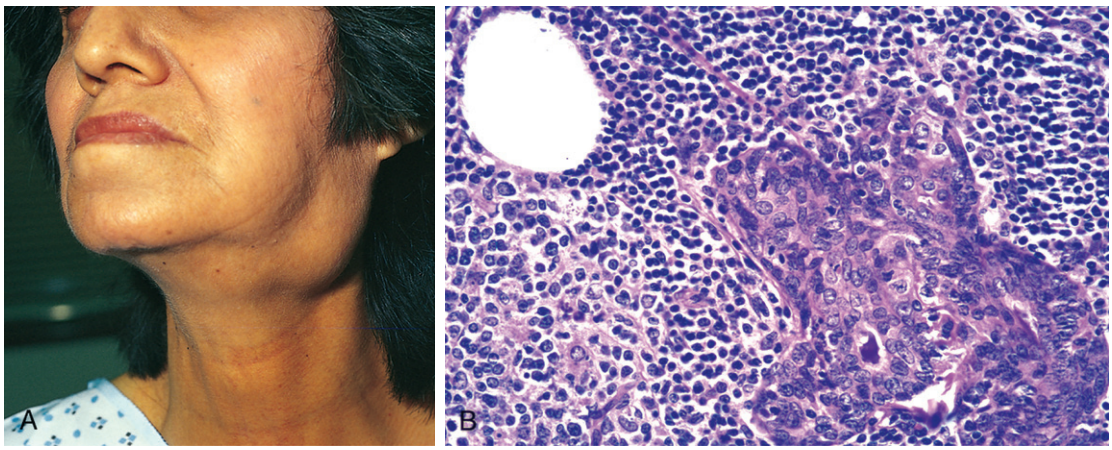


Figure 4–20 Sjögren syndrome. **A,** Enlargement of the salivary gland. **B,** Histopathologic findings include intense lymphocytic and plasma cell infiltration with ductal epithelial hyperplasia.
(A, Courtesy of Dr. Richard Sontheimer, Department of Dermatology, University of Texas Southwestern Medical School, Dallas, Texas. B, Courtesy of Dr. Dennis Burns, Department of Pathology, University of Texas Southwestern Medical School, Dallas, Texas.)

MORPHOLOGY

Lacrimal and salivary glands are the primary targets, but other secretory glands, including those in the nasopharynx, upper airway, and vagina, also may be involved. Histologic examination shows an intense lymphocyte (primarily activated CD4+ T cells) and plasma cell infiltrate, occasionally forming lymphoid follicles with germinal centers. There is associated destruction of the native architecture (Fig. 4–20).

Lacrimal gland destruction results in a lack of tears, leading to drying of the corneal epithelium, with subsequent inflammation, erosion, and ulceration **(keratoconjunctivitis).** Similar changes may occur in the oral mucosa as a result of loss of salivary gland output, giving rise to mucosal atrophy, with inflammatory fissuring and ulceration **(xerostomia).** Dryness and crusting of the nose may lead to ulcerations and even perforation of the nasal septum. When the respiratory passages are involved, secondary laryngitis, bronchitis, and pneumonitis may appear. Approximately 25% of the patients (especially those with anti-SS-A antibodies) acquire extraglandular disease involving the CNS, skin, kidneys, and muscles. Renal lesions take the form of mild interstitial nephritis associated with tubular transport defects; unlike in SLE, glomerulonephritis is rare.

Clinical Course

Approximately 90% of Sjögren syndrome cases occur in women between the ages of 35 and 45 years. Patients present with dry mouth, lack of tears, and the resultant complications described above. Salivary glands are often enlarged as a result of lymphocytic infiltrates (Fig. 4–20). Extraglandular manifestations include synovitis, pulmonary fibrosis, and peripheral neuropathy. About 60% of Sjögren patients have another accompanying autoimmune disorder such as RA. Notably, there is a 40-fold increased risk for developing a non-Hodgkin B cell lymphoma, arising in the setting of the initial robust polyclonal B cell proliferation. These so-called marginal zone lymphomas are discussed in Chapter 11.

SUMMARY

Sjögren Syndrome

- Sjögren syndrome is an inflammatory disease that affects primarily the salivary and lacrimal glands, causing dryness of the mouth and eyes.
- The disease is believed to be caused by an autoimmune T cell reaction against one or more unknown self antigens expressed in these glands, or immune reactions against the antigens of a virus that infects the tissues.

Systemic Sclerosis (Scleroderma)

Systemic sclerosis (SS) is an immunologic disorder characterized by excessive fibrosis in multiple tissues, obliterative vascular disease, and evidence of autoimmunity, mainly the production of multiple autoantibodies. It is commonly called scleroderma because the skin is a major target, but this disorder is better labeled "systemic" because lesions are present throughout the body. Cutaneous involvement is the usual presenting manifestation and eventually appears in approximately 95% of cases, but it is the visceral involvement—of the gastrointestinal tract, lungs, kidneys, heart, and skeletal muscles—that is responsible for most of the related morbidity and mortality.

SS can be classified into two groups on the basis of its clinical course:

- *Diffuse scleroderma,* characterized by initial widespread skin involvement, with rapid progression and early visceral involvement
- *Limited scleroderma,* with relatively mild skin involvement, often confined to the fingers and face. Involvement of the viscera occurs late, so the disease in these patients generally has a fairly benign course. This clinical presentation is also called the CREST syndrome because of its frequent features of calcinosis, Raynaud phenomenon, esophageal dysmotility, sclerodactyly, and telangiectasia.

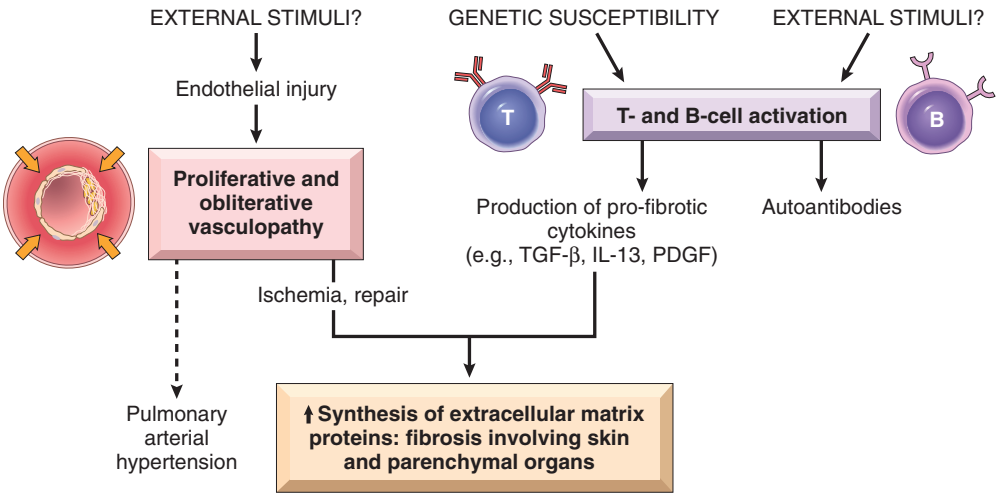


Figure 4–21 A model for the pathogenesis of systemic sclerosis. Unknown external stimuli cause vascular abnormalities and immune activation in genetically susceptible individuals, and both contribute to the excessive fibrosis.

PATHOGENESIS

The cause of the disease is not known, but genetic and environmental factors probably contribute. A postulated sequence of events is the following (Fig. 4–21).

- **Injury to endothelial cells** of small arteries by unknown mechanisms leads to endothelial activation, increased expression of adhesion molecules, and migration of activated T cells into the perivascular tissues. The local T cell reaction may cause further activation and injury to endothelial cells.
- **T cells** respond to some self antigen and produce cytokines. It has been suggested that the dominant T cells are T_H2 cells, and their cytokines induce alternative macrophage activation and collagen deposition. The activated T cells and macrophages produce **cytokines that activate fibroblasts and stimulate collagen production,** resulting in fibrosis. These cytokines include TGF-β, IL-13, platelet-derived growth factor (PDGF), and others.
- Repeated bouts of endothelial damage followed by platelet aggregation lead to **endothelial proliferation and intimal fibrosis,** which, together with periadventitial fibrosis, narrow the small vessels, with eventual **ischemic injury.** The subsequent repair reaction may lead to more fibrosis, thus setting up a self-perpetuating cycle.
- **B cell activation** also occurs, as indicated by the presence of hypergammaglobulinemia and ANAs. Although there is no evidence that humoral immunity plays a significant role in the pathogenesis of SS, two of the ANAs are virtually unique to this disease and are therefore useful in diagnosis (Table 4–10). One of these, directed **against DNA topoisomerase I** (anti-Scl 70), is highly specific; it is present in as many as 70% of patients with diffuse scleroderma (and in less than 1% of patients with other connective tissue diseases) and is a marker for the development of more aggressive disease with pulmonary fibrosis and peripheral vascular changes. The other ANA is an **anticentromere antibody,** found in as many as 90% of patients with limited scleroderma (i.e., the CREST syndrome); it indicates a relatively benign course.

MORPHOLOGY

Virtually any organ may be affected in SS, but the most prominent changes are found in the skin, musculoskeletal system, gastrointestinal tract, lungs, kidneys, and heart.

Skin. The vast majority of patients have diffuse, sclerotic atrophy of the skin, usually beginning in the fingers and distal regions of the upper extremities and extending proximally to involve the upper arms, shoulders, neck, and face. In the early stages, affected skin areas are somewhat edematous and have a doughy consistency. Histopathologic findings include edema and perivascular infiltrates containing CD4+ T cells. Capillaries and small arteries (as large as 500 μm in diameter) may show thickening of the basal lamina, endothelial cell damage, and partial occlusion. With progression, the edematous phase is replaced by progressive fibrosis of the dermis, which becomes tightly bound to the subcutaneous structures. There is marked increase of compact collagen in the dermis along with thinning of the epidermis, atrophy of the dermal appendages, and hyaline thickening of the walls of dermal arterioles and capillaries (Fig. 4–22, *A, B*). Focal and sometimes diffuse subcutaneous calcifications may develop, especially in patients with the CREST syndrome. In advanced stages, the fingers take on a tapered, clawlike appearance with limitation of motion in the joints (Fig. 4–22, *C*), and the face becomes a drawn mask. Loss of blood supply may lead to cutaneous ulcerations and to atrophic changes in the terminal phalanges, including autoamputation.

Gastrointestinal Tract. The gastrointestinal tract is affected in approximately 90% of patients. Progressive atrophy and collagenous fibrous replacement of the muscularis may develop at any level of the gut but are most severe in the esophagus, with the lower two thirds often demonstrating an inflexibility not unlike that typical of a rubber hose. The associated dysfunction of the lower esophageal sphincter gives rise to gastroesophageal reflux and its complications, including Barrett metaplasia (Chapter 14) and strictures. The mucosa is thinned and may be ulcerated, and there is excessive collagenization of the lamina propria and submucosa.

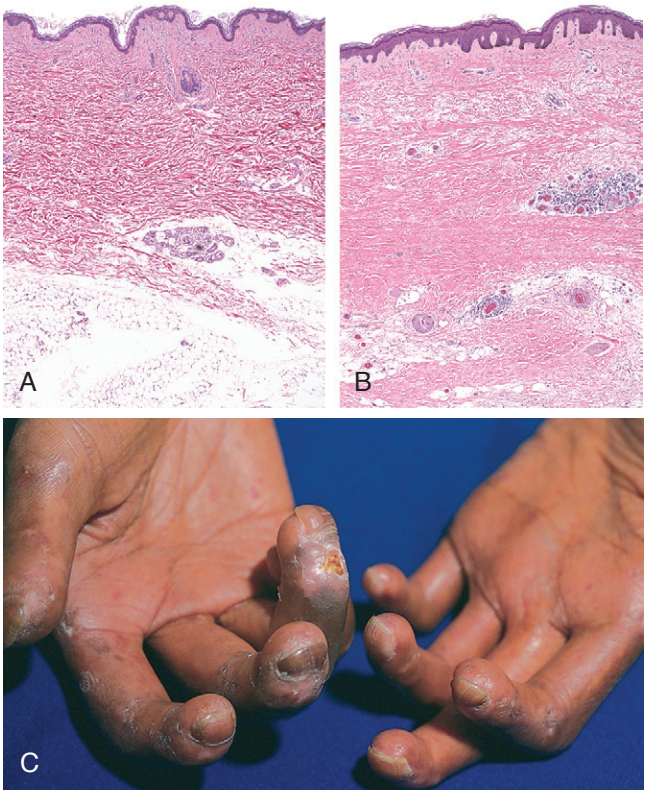


Figure 4–22 Systemic sclerosis. **A,** Normal skin. **B,** Extensive deposition of dense collagen in the dermis. **C,** The extensive subcutaneous fibrosis has virtually immobilized the fingers, creating a clawlike flexion deformity. Loss of blood supply has led to cutaneous ulcerations.
(A–C, Courtesy of Dr. Richard Sontheimer, Department of Dermatology, University of Texas Southwestern Medical School, Dallas, Texas.)

Loss of villi and microvilli in the small bowel is the anatomic basis for the malabsorption syndrome sometimes encountered in affected patients.

Musculoskeletal System. Synovial hyperplasia and inflammation are common in the early stages; fibrosis later ensues. Although these changes are reminiscent of RA, joint destruction is not common in SS. In a small subset of patients (approximately 10%), inflammatory myositis indistinguishable from polymyositis may develop.

Lungs. The lungs are affected in more than 50% of patients; lung involvement may manifest as pulmonary hypertension and/or interstitial fibrosis. Pulmonary vasospasm from pulmonary vascular endothelial dysfunction is considered important in the pathogenesis of pulmonary hypertension. Pulmonary fibrosis, when present, is indistinguishable from that seen in idiopathic pulmonary fibrosis (Chapter 12).

Kidneys. Renal abnormalities occur in two thirds of patients with SS, most typically associated with thickening of the vessel walls of interlobular arteries (150 to 500 μm in diameter). These show intimal cell proliferation with deposition of various glycoproteins and acid mucopolysaccharides. Although similar to the changes seen in malignant hypertension, the alterations in SS are restricted to vessels 150 to 500 μm in diameter and are not always associated with hypertension. Hypertension does occur in 30% of the patients and, in 20% of those patients, takes an ominously malignant course (malignant hypertension). In hypertensive patients, vascular alterations are more pronounced and are often associated with fibrinoid necrosis involving the arterioles together with thrombosis and infarction. Such patients often die of renal failure, accounting for about half of the deaths attributable to SS. There are no specific glomerular changes.

Heart. Patchy myocardial fibrosis, along with thickening of intramyocardial arterioles, occurs in one-third of the patients; this is putatively caused by microvascular injury and resultant ischemia (so-called cardiac Raynaud). Because of the changes in the lung, right ventricular hypertrophy and failure (cor pulmonale) are frequent.

Clinical Course

SS affects women three times more often than men, with a peak incidence in the 50- to 60-year age group. There is a substantial overlap in presentation between SS and RA, SLE, and dermatomyositis (see later); *the distinctive feature of SS is the striking cutaneous involvement*. Almost all patients exhibit Raynaud phenomenon, a vascular disorder characterized by reversible vasospasm of the arteries. Typically the hands turn white on exposure to cold, reflecting vasospasm, followed by change to blue as ischemia and cyanosis supervene. Finally, the color changes to red as reactive vasodilation occurs. Progressive collagen deposition in the skin leads to atrophy of the hands, with increasing stiffness and eventually complete immobilization of the joints. Difficulty in swallowing results from esophageal fibrosis and resultant hypomotility. Eventually, destruction of the esophageal wall leads to atony and dilation. Malabsorption may appear if the submucosal and muscular atrophy and fibrosis involve the small intestine. Dyspnea and chronic cough reflect the pulmonary changes; with advanced lung involvement, *secondary pulmonary hypertension* may develop, leading to right-sided cardiac failure. Renal functional impairment secondary to both the advance of SS and the concomitant malignant hypertension is frequently marked.

The clinical course for diffuse SS is difficult to predict. In most patients the disease pursues a steady, slow, downhill course over the span of many years, although in the absence of renal involvement, life span may be normal. The overall 10-year survival rate ranges from 35% to 70%. The chances of survival are significantly better for patients with localized scleroderma than for those with the usual diffuse progressive disease. Limited scleroderma, or CREST syndrome, frequently has Raynaud phenomenon as its presenting feature. It is associated with limited skin involvement confined to the fingers and face, and these two features may be present for decades before the appearance of visceral lesions.

SUMMARY

Systemic Sclerosis

- SS (commonly called scleroderma) is characterized by progressive fibrosis involving the skin, gastrointestinal tract, and other tissues.

- Fibrosis may be the result of activation of fibroblasts by cytokines produced by T cells, but what triggers T cell responses is unknown.
- Endothelial injury and microvascular disease are commonly present in the lesions of SS, causing chronic ischemia, but the pathogenesis of vascular injury is not known.

Inflammatory Myopathies

Inflammatory myopathies make up a heterogeneous group of rare disorders characterized by immune-mediated muscle injury and inflammation. Based on the clinical, morphologic, and immunologic features, three disorders—polymyositis, dermatomyositis, and inclusion body myositis—have been described. These are discussed in Chapter 21.

Mixed Connective Tissue Disease

The term *mixed connective tissue disease* refers to a spectrum of pathologic processes in patients who present with clinical features suggestive of SLE, polymyositis, or SS; they also have high titers of antibodies to an RNP antigen called U1RNP. Two other features of mixed connective tissue disease are the paucity of renal disease and an extremely good response to corticosteroids, both of which suggest a favorable long-term prognosis.

Mixed connective tissue disease may manifest as arthritis, swelling of the hands, Raynaud phenomenon, esophageal dysmotility, myositis, leukopenia and anemia, fever, lymphadenopathy, and/or hypergammaglobulinemia. Because of these overlapping features, it is not entirely clear whether mixed connective tissue disease constitutes a distinct clinical entity or if such disorders represent heterogeneous subsets of SLE, systemic sclerosis, and polymyositis; most authorities do not consider it to be a specific entity.

Polyarteritis Nodosa and Other Vasculitides

Polyarteritis nodosa belongs to a group of diseases characterized by necrotizing inflammation of the walls of blood vessels, most likely caused by deposition of immune complexes. The general term *noninfectious necrotizing vasculitis* differentiates these conditions from those attributable to direct vessel infection (e.g., an abscess) and serves to emphasize that any type of vessel may be involved—arteries, arterioles, veins, or capillaries. A detailed classification and description of vasculitides are presented in Chapter 9.

IgG4-Related Disease

IgG4-related disease (IgG4-RD) is a newly recognized fibroinflammatory condition characterized by a tendency to form tumor-like lesions in several organs. The disorder is often, but not always, associated with elevated serum IgG4 concentrations. However, increased numbers of IgG4-producing plasma cells (or an increased IgG4 to total IgG ratio) in tissue are a sine qua non of this disorder. Although recognized only recently when extra-pancreatic manifestations were identified in patients with autoimmune pancreatitis, IgG4-RD has now been described in virtually every organ system: the biliary tree, salivary glands, periorbital tissues, kidneys, lungs, lymph nodes, meninges, aorta, breast, prostate, thyroid, pericardium, and skin. The histopathologic features bear striking similarities across organs, regardless of the site of disease. These include a mixed infiltrate of lymphoid cells (T cells, B cells, plasma cells), storiform fibrosis, obliterative phlebitis, and mild to moderate tissue eosinophilia. B cells are typically organized in germinal centers but the T cells—the predominant cell type—are distributed diffusely throughout the lesion. The ratio of IgG4 to IgG-bearing plasma cells (determined by semiquantitative immunohistochemistry) is typically equal to or greater than 50%.

Many medical conditions long viewed as confined to single organs are part of the IgG4-RD spectrum. These include Mikulicz syndrome (enlargement and fibrosis of salivary and lacrimal glands), Riedel thyroiditis, idiopathic retroperitoneal fibrosis, autoimmune pancreatitis, and inflammatory pseudotumors of the orbit, lungs, and kidneys, to name a few.

The role of IgG4 in the pathogenesis of this condition is not fully understood. However, the key role of B cells is supported by initial studies in which depletion of B cells by anti–B cell reagents such as Rituximab provided clinical benefit. It is unclear if the disease is truly autoimmune in nature, and no target autoantigens have been identified.

REJECTION OF TRANSPLANTS

The major barrier to transplantation of organs from one individual to another of the same species (called allografts) is immunologic rejection of the transplanted tissue. Rejection is a complex phenomenon involving both cell- and antibody-mediated reactions that destroy the graft. The key to successful transplantation has been the development of therapies that prevent or minimize rejection. Discussed next is how grafts are recognized as foreign and how they are rejected.

Immune Recognition of Allografts

Rejection of allografts is a response mainly to MHC molecules, which are so polymorphic that most individuals in an outbred population differ in at least some of the MHC molecules they express (except, of course, for identical twins). There are two main mechanisms by which the host immune system recognizes and responds to the MHC molecules on the graft (Fig. 4–23):

- *Direct recognition.* Host T cells directly recognize the allogeneic (foreign) MHC molecules that are expressed on graft cells. Direct recognition of foreign MHC seems to violate the rule of MHC restriction, which states that in every individual, all of the T cells are educated to recognize foreign antigens displayed by only that individual's MHC molecules. It is postulated that allogeneic MHC molecules (with any bound peptides) structurally mimic

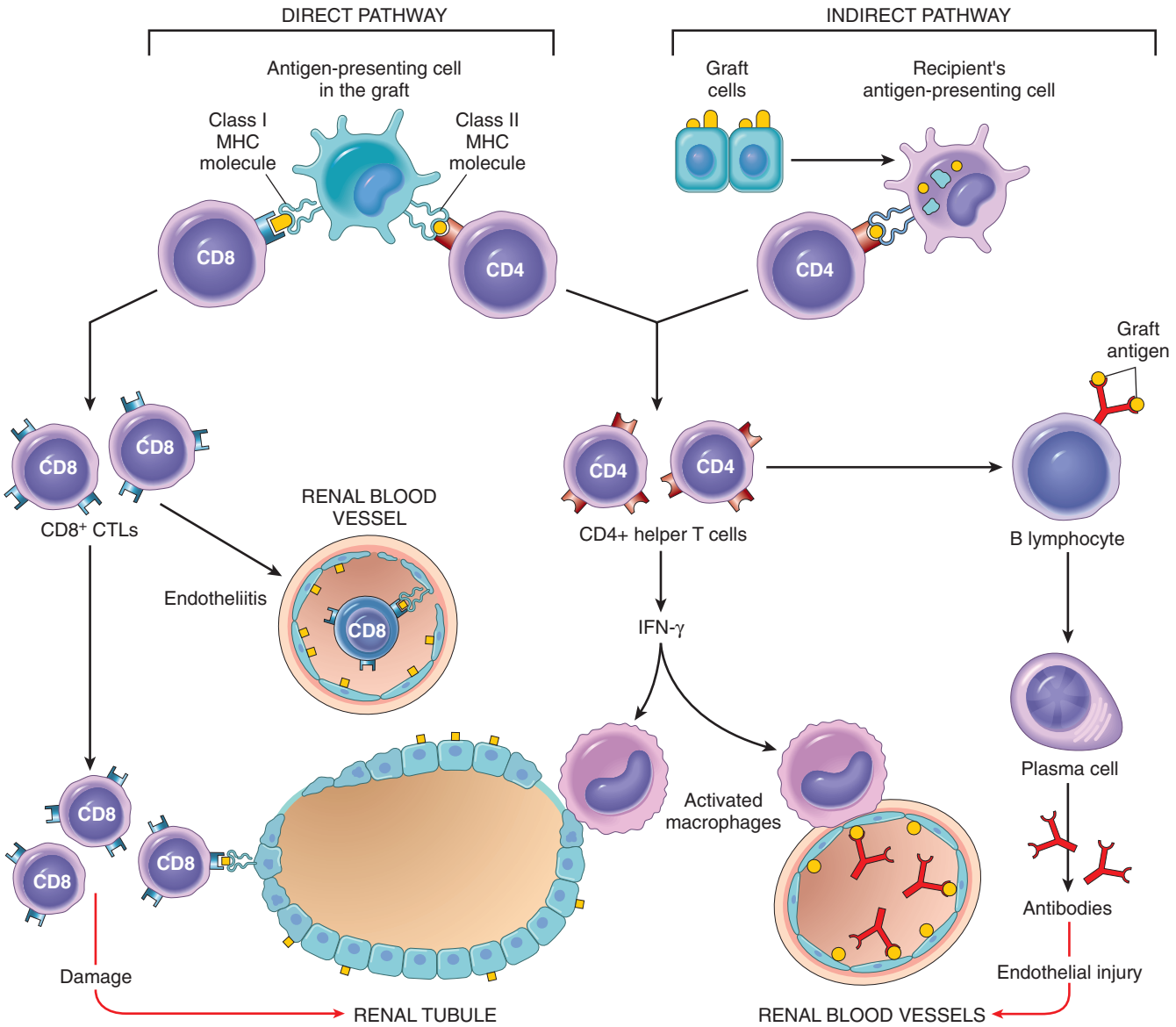


Figure 4–23 Recognition and rejection of allografts. In the direct pathway, donor class I and class II MHC antigens on antigen-presenting cells (APCs) in the graft (along with costimulators, *not shown*) are recognized by host CD8+ cytotoxic T cells and CD4+ helper T cells, respectively. CD4+ cells proliferate and produce cytokines (e.g., IFN-γ), which induce tissue damage by a local delayed-type hypersensitivity reaction. CD8+ T cells responding to graft antigens differentiate into CTLs that kill graft cells. In the indirect pathway, graft antigens are displayed by host APCs and activate CD4+ T cells, which damage the graft by a local delayed-type hypersensitivity reaction and stimulate B lymphocytes to produce antibodies. IFN-γ, interferon-γ; MHC, major histocompatibility complex.

self MHC and foreign peptide, and so direct recognition of the allogeneic MHC is essentially an immunologic cross-reaction. Because DCs in the graft express high levels of MHC as well as costimulatory molecules, they are believed to be the major culprits contributing to direct recognition. The most important consequence of direct recognition is the activation of host CD8+ T cells that recognize class I MHC (HLA-A, -B) molecules in the graft. These T cells differentiate into CTLs, which kill the cells in the graft. Host CD4+ helper T cells may be triggered into proliferation and cytokine production by recognition of donor class II MHC (HLA-D) molecules and drive an inflammatory response.

- *Indirect recognition.* In this pathway, host CD4+ T cells recognize donor MHC molecules after these molecules are picked up, processed, and presented by the host's own APCs. This sequence is similar to the physiologic processing and presentation of other foreign (e.g., microbial) antigens. The activated CD4+ T cells then recognize APCs displaying graft antigens and secrete cytokines that induce inflammation and damage the graft. The indirect pathway is also involved in the production of antibodies against graft alloantigens; if these antigens are proteins, they are picked up by host B cells, and peptides are presented to helper T cells, which then stimulate antibody responses.

Effector Mechanisms of Graft Rejection

Both T cells and antibodies reactive with the graft are involved in the rejection of most solid-organ allografts (Fig. 4–23).

T Cell–Mediated Rejection

CTLs kill cells in the grafted tissue, causing parenchymal and endothelial cell death (the latter resulting in thrombosis and graft ischemia). Cytokine-secreting CD4+ T cells trigger inflammatory reactions resembling DTH in the tissues and blood vessels, with local accumulation of mononuclear cells (lymphocytes and macrophages). Activated microphages can injure graft cells and vasculature. The microvascular injury also results in tissue ischemia, which contributes to graft destruction.

Antibody-Mediated Rejection

Although T cells are of paramount importance in allograft rejection, antibodies also mediate some forms of rejection. Alloantibodies directed against graft MHC molecules and other alloantigens bind to the graft endothelium and cause vascular injury through complement activation and recruitment of leukocytes. Superimposed on the resulting endothelial damage and dysfunction is thrombosis, adding further ischemic insult to the injury.

Hyperacute rejection is a special form of rejection occurring if *pre-formed anti-donor antibodies* are present in the circulation of the host before transplantation. This may happen in multiparous women who have anti-HLA antibodies against paternal antigens encountered during pregnancy, or in individuals exposed to foreign HLA (on platelets or leukocytes) from previous blood transfusions. Obviously, such antibodies also may be present in a patient who has previously rejected an organ transplant. Subsequent transplantation in such patients will result in immediate rejection (within minutes to hours) because the circulating antibodies rapidly bind to the endothelium of the grafted organ, with resultant complement activation and vascular thrombosis. With the current practice of screening potential recipients for pre-formed anti-HLA antibodies and cross-matching (testing recipients for the presence of antibodies directed against the donor's lymphocytes), hyperacute rejection occurs in less than 0.4% of transplant recipients.

MORPHOLOGY

On the basis of the time course and morphology of rejection reactions, they have been classified as hyperacute, acute, and chronic (Fig. 4–24). This classification is helpful for understanding the mechanism of rejection, because each pattern is caused by a different type of dominant immunologic reaction. The morphology of these patterns is described in the context of renal transplants; however, similar changes are encountered in other vascularized organ transplants.

Hyperacute Rejection

Hyperacute rejection occurs within minutes to a few hours after transplantation in a presensitized host and typically is recognized by the surgeon just after the vascular anastomosis is completed. In contrast with a nonrejecting kidney graft, which regains a normal pink color and tissue turgor and promptly excretes urine, a hyperacutely rejecting kidney rapidly becomes cyanotic, mottled, and flaccid and may excrete only a few drops of bloody fluid. The histologic picture is characterized by widespread acute arteritis and arteriolitis, vessel thrombosis, and ischemic necrosis, all resulting from the binding of preformed antibodies to graft endothelium. Virtually all arterioles and arteries exhibit characteristic acute fibrinoid necrosis of their walls, with narrowing or complete occlusion of the lumens by precipitated fibrin and cellular debris (Fig. 4–24, *A*).

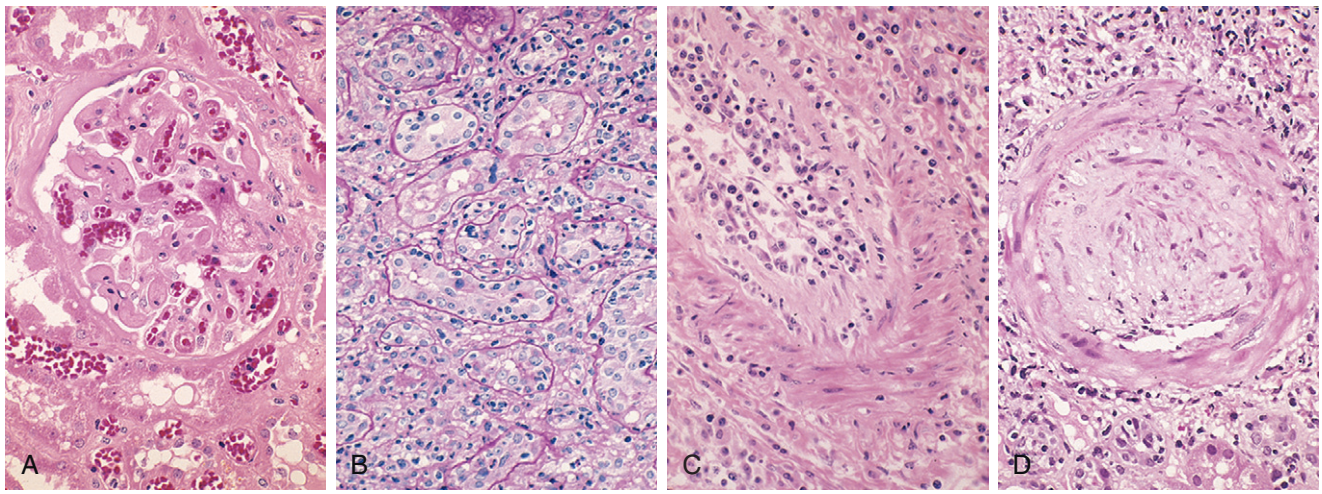


Figure 4–24 Morphologic patterns of graft rejection. **A,** Hyperacute rejection of a kidney allograft associated with endothelial damage and thrombi in a glomerulus. **B,** Acute cellular rejection of a kidney allograft with inflammatory cells in the interstitium and between epithelial cells of the tubules. **C,** Acute humoral rejection of a kidney allograft (rejection vasculitis) with inflammatory cells and proliferating smooth muscle cells in the intima. **D,** Chronic rejection in a kidney allograft with graft arteriosclerosis. The arterial lumen is replaced by an accumulation of smooth muscle cells and connective tissue in the intima.

(A–D, Courtesy of Dr. Helmut Rennke, Department of Pathology, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts.)

Acute Rejection

Acute rejection may occur within days to weeks of transplantation in a nonimmunosuppressed host or may appear months or even years later, even in the presence of adequate immunosuppression. Acute rejection is caused by both cellular and humoral immune mechanisms, and in any one patient, one or the other may predominate, or both may be present. On histologic examination, cellular rejection is marked by an interstitial mononuclear cell infiltrate with associated edema and parenchymal injury, whereas humoral rejection is associated with vasculitis.

Acute cellular rejection most commonly is seen within the first months after transplantation and typically is accompanied by clinical signs of renal failure. Histologic examination usually shows extensive interstitial CD4+ and CD8+ T cell infiltration with edema and mild interstitial hemorrhage (Fig. 4–24, *B*). Glomerular and peritubular capillaries contain large numbers of mononuclear cells, which also may invade the tubules, leading to focal tubular necrosis. In addition to tubular injury, CD8+ T cells also may injure the endothelium, causing an endothelitis. Cyclosporine (a widely used immunosuppressive agent) is also nephrotoxic and induces so-called arteriolar hyaline deposits. Renal biopsy is used to distinguish rejection from drug toxicity. Accurate recognition of cellular rejection is important, because patients typically respond promptly to increased immunosuppressive therapy.

Acute humoral rejection (rejection vasculitis) caused by antidonor antibodies also may participate in acute graft rejection. The histologic lesions may take the form of necrotizing vasculitis with endothelial cell necrosis; neutrophilic infiltration; deposition of antibody, complement, and fibrin; and thrombosis. Such lesions may be associated with ischemic necrosis of the renal parenchyma. Somewhat older subacute lesions are characterized by marked thickening of the intima by proliferating fibroblasts, myocytes, and foamy macrophages (Fig. 4–24, *C*). The resultant narrowing of the arterioles may cause infarction or renal cortical atrophy. The proliferative vascular lesions mimic arteriosclerotic thickening and are believed to be caused by cytokines that stimulate proliferation of vascular smooth muscle cells. Local deposition of complement breakdown products (specifically C4d) is used to detect antibody-mediated rejection of kidney allografts.

Chronic Rejection

Patients present with chronic rejection late after transplantation (months to years) with a progressive rise in serum creatinine levels (an index of renal function) over a period of 4 to 6 months. Chronic rejection is dominated by vascular changes, interstitial fibrosis, and loss of renal parenchyma; there are typically only mild or no ongoing cellular parenchymal infiltrates. The vascular changes occur predominantly in the arteries and arterioles, which exhibit intimal smooth muscle cell proliferation and extracellular matrix synthesis (Fig. 4–24, *D*). These lesions ultimately compromise vascular perfusion and result in renal ischemia manifested by loss or hyalinization of glomeruli, interstitial fibrosis, and tubular atrophy. The vascular lesion may be caused by cytokines released by activated T cells that act on the cells of the vascular wall, and it may be the end stage of the proliferative arteritis described earlier.

SUMMARY

Recognition and Rejection of Organ Transplants (Allografts)

- The graft rejection response is initiated mainly by host T cells that recognize the foreign HLA antigens of the graft, either directly (on APCs in the graft) or indirectly (after uptake and presentation by host APCs).
- Types and mechanisms of rejection comprise the following:
  - *Hyperacute rejection*: Pre-formed antidonor antibodies bind to graft endothelium immediately after transplantation, leading to thrombosis, ischemic damage, and rapid graft failure.
  - *Acute cellular rejection*: T cells destroy graft parenchyma (and vessels) by cytotoxicity and inflammatory reactions.
  - *Acute humoral rejection*: Antibodies damage graft vasculature.
  - *Chronic rejection*: Dominated by arteriosclerosis, this type is probably caused by T cell reaction and secretion of cytokines that induce proliferation of vascular smooth muscle cells, associated with parenchymal fibrosis.

Methods of Improving Graft Survival

Because HLA molecules are the major targets in transplant rejection, better matching of the donor and the recipient improves graft survival. The benefits of HLA matching are most dramatic for living related donor kidney transplants, and survival improves with increasing number of loci matched. However, as drugs for immunosuppression have improved, HLA matching is not even attempted in some situations, such as heart, lung, liver, and islet transplantation; in such instances, the recipient often needs a transplant urgently and other considerations, such as anatomic (size) compatibility, are of greater practical importance.

Immunosuppression of the recipient is a practical necessity in all organ transplantation except in the case of identical twins. At present, drugs such as cyclosporine, the related FK506, mofetil mycophenolate (MMF), rapamycin, azathioprine, corticosteroids, antilymphocyte globulin, and monoclonal antibodies (e.g., monoclonal anti-CD3) are used. Cyclosporine and FK506 suppress T cell–mediated immunity by inhibiting transcription of cytokine genes, in particular, the gene for IL-2. Although immunosuppression has made transplantation of many organs feasible, there is still a price to be paid. Global immunosuppression results in increased susceptibility to opportunistic fungal, viral, and other infections. These patients are also at increased risk for developing Epstein-Barr virus (EBV)–induced lymphomas, human papillomavirus–induced squamous cell carcinomas, and Kaposi sarcoma. To circumvent the untoward effects of immunosuppression, much effort is devoted to trying to induce donor-specific tolerance in host T cells. One strategy being pursued is to prevent host T cells from receiving costimulatory signals from donor DCs during the initial phase of sensitization. This can be accomplished by administration of agents to interrupt the interaction between the B7 molecules on the DCs of the graft with the CD28 receptors on host T cells.

This will interrupt the second signal for T cell activation and either induce apoptosis or render the T cells functionally unresponsive. The improvement in immunosuppressive therapy has led to improving survival of grafts and acute rejection is becoming less of a concern, especially for kidney and heart grafts. Chronic rejection remains a serious problem, however, especially because it responds much less effectively than does acute rejection to available immunosuppressive agents.

Transplantation of Hematopoietic Stem Cells

Hematopoietic stem cell (HSC) transplantation is used as therapy for hematopoietic and some nonhematopoietic malignancies, aplastic anemias, and certain inherited disorders, particularly immune deficiency states and severe forms of thalassemia. HSCs historically were obtained solely from donor bone marrow, but are now increasingly harvested from the peripheral blood after mobilization by administration of hematopoietic growth factors, or from the umbilical cord blood of newborns, a readily available rich source of HSCs. The recipient receives chemotherapy and/or irradiation to destroy malignant cells (e.g., in leukemia) and to create a graft bed; then, HSCs are infused into the peripheral blood, from which they home to their bone marrow niches. Rejection of allogeneic HSC transplants seems to be mediated by some combination of host T cells and NK cells that are resistant to radiation therapy and chemotherapy. Two major problems complicate this form of transplantation: graft-versus-host disease and immune deficiency.

Graft-Versus-Host Disease (GVHD). This occurs when immunologically competent T cells (or their precursors) are transplanted into recipients who are immunologically compromised. Although GVHD happens most commonly in the setting of allogeneic HSC transplantation (usually involving minor histocompatibility mismatches between donor and recipient), it also may occur after transplantation of solid organs rich in lymphoid cells (e.g., the liver) or after transfusion of nonirradiated blood. On receiving allogeneic HSCs, an immunologically compromised host cannot reject the graft, but T cells present in the donor graft perceive the recipient's tissue as "foreign" and react against it. This results in the activation of both CD4+ and CD8+ T cells, ultimately causing inflammation and killing host cells.

- *Acute GVHD* (occurring days to weeks after transplantation) causes epithelial cell necrosis in three principal target organs: liver, skin, and gut. Destruction of small bile ducts gives rise to jaundice, and mucosal ulceration of the gut results in bloody diarrhea. Cutaneous involvement is manifested by a generalized rash.
- *Chronic GVHD* may follow the acute syndrome or may occur insidiously. The patients develop skin lesions resembling those of SS (discussed earlier) and manifestations mimicking other autoimmune disorders.

GVHD is a potentially lethal complication that can be minimized but not eliminated by HLA matching. As another potential solution, donor T cells can be depleted before marrow transplant. This protocol has proved to be a mixed blessing: The risk of GVHD is reduced, but the incidence of graft failure and (in those with the disease) the recurrence of leukemia increase. It seems that the multifunctional T cells not only mediate GVHD but also are required for the efficient engraftment of the transplanted HSCs and elimination of leukemia cells (so-called graft-versus-leukemia effect).

Immune Deficiencies. These are often of prolonged duration in recipients of HSC transplants. Among the many reasons for this impairment is the slow reconstitution of the recipient's adaptive immune system, which is destroyed or suppressed to allow the graft to take and requires many months to recover. During this vulnerable period, recipients are susceptible to a variety of infections, mostly viral, such as cytomegalovirus (CMV) and EBV infections.

IMMUNE DEFICIENCY DISEASES

Immune deficiency diseases may be caused by inherited defects affecting immune system development, or they may result from secondary effects of other diseases (e.g., infection, malnutrition, aging, immunosuppression, autoimmunity, or chemotherapy). *Clinically, patients with immune deficiency present with increased susceptibility to infections as well as to certain forms of cancer.* The type of infections in a given patient depends largely on the component of the immune system that is affected. Patients with defects in immunoglobulin, complement, or phagocytic cells typically suffer from recurrent infections with pyogenic bacteria, whereas those with defects in cell-mediated immunity are prone to infections caused by viruses, fungi, and intracellular bacteria. Discussed next are some of the more important primary (congenital) immune deficiencies, followed by a detailed description of the acquired immunodeficiency syndrome (AIDS), the most devastating example of secondary (acquired) immune deficiency.

Primary (Congenital) Immune Deficiencies

Primary immune deficiency states are fortunately rare, but their study has nevertheless contributed greatly to the current understanding of the development and function of the immune system. Most primary immune deficiency diseases are genetically determined and affect either adaptive immunity (i.e., humoral or cellular) or innate host defense mechanisms, including complement proteins and cells such as phagocytes and NK cells. Defects in adaptive immunity are often subclassified on the basis of the primary component involved (i.e., B cells or T cells, or both); however, because of the interactions between T and B lymphocytes, these distinctions are not clear-cut. For instance, T cell defects frequently lead to impaired antibody synthesis, and hence isolated deficiencies of T cells may be indistinguishable from combined deficiencies of T and B cells. Most primary immune deficiencies come to attention early in life (between the ages of 6 months and 2 years), usually because the affected infants are susceptible to recurrent infections. One of the most impressive accomplishments of modern molecular biology has been the identification of the genetic basis for many primary immune deficiencies (Fig. 4–25), laying the foundation for future gene replacement therapy.

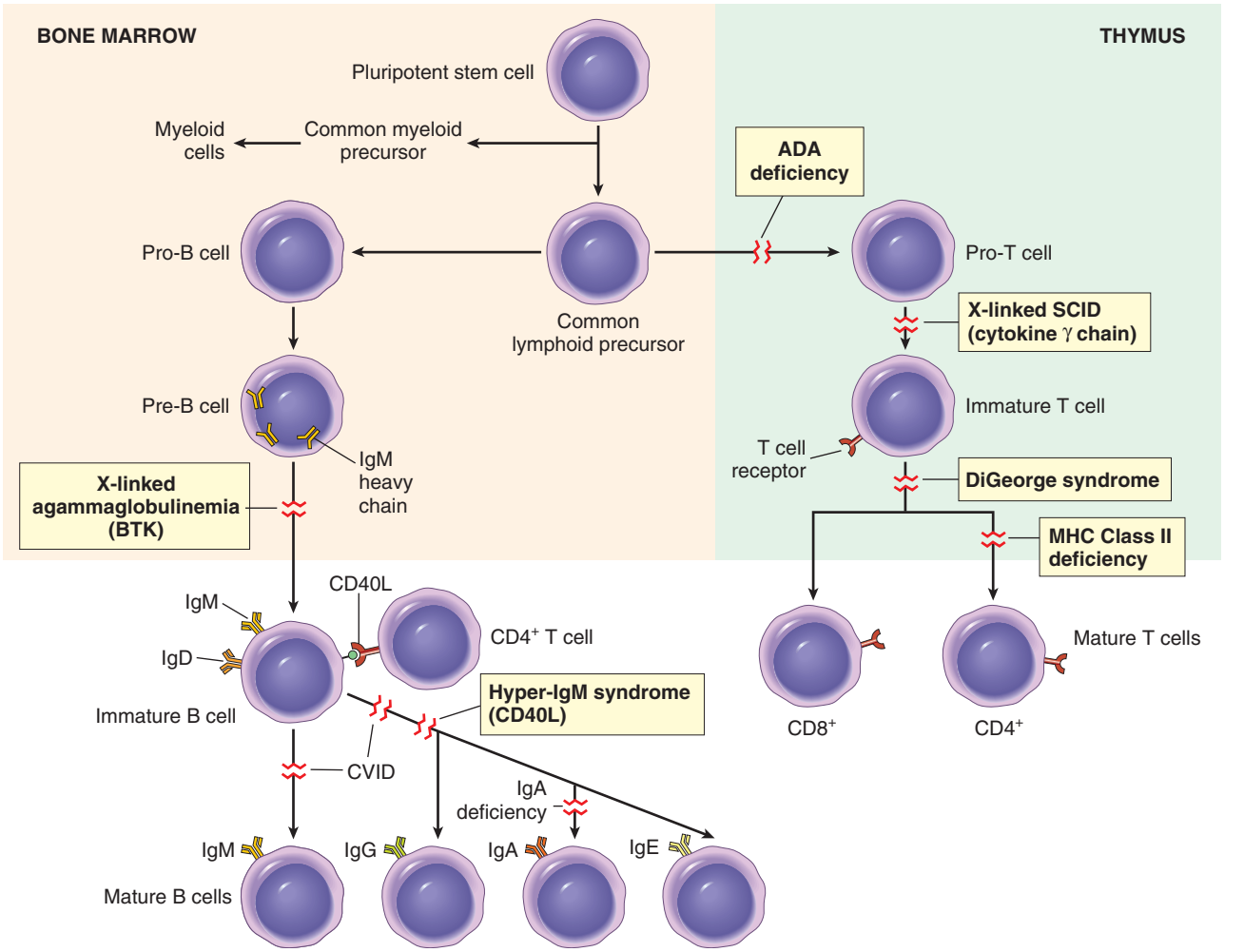


Figure 4–25 Primary immune deficiency diseases. Shown are the principal pathways of lymphocyte development and the blocks in these pathways in selected primary immune deficiency diseases. The affected genes are indicated in *parentheses* for some of the disorders. ADA, adenosine deaminase; CD40L, CD40 ligand (also known as CD154); CVID, common variable immunodeficiency; SCID, severe combined immunodeficiency.

X-Linked Agammaglobulinemia: Bruton Disease

X-linked agammaglobulinemia (XLA), or Bruton disease, is characterized by the failure of pre-B cells to differentiate into B cells and, as the name implies, there is a resultant absence of antibodies (gamma globulin) in the blood. It occurs at a frequency of about 1 in 100,000 male infants. During normal B cell maturation, immunoglobulin heavy chain genes are rearranged first, followed by light chain rearrangement. At each stage, signals are received from the expressed components of the antigen receptor that drive maturation to the next stage; these signals act as quality controls, to ensure that the correct receptor proteins are being produced. In XLA, B cell maturation stops after the initial heavy chain gene rearrangement because of mutations in a tyrosine kinase that is associated with the pre-B cell receptor and is involved in pre-B cell signal transduction. This kinase is called Bruton tyrosine kinase (BTK). When it is nonfunctional, the pre-B cell receptor cannot signal the cells to proceed along the maturation pathway. As a result, immunoglobulin light chains are not produced, and the complete immunoglobulin molecule containing heavy and light chains cannot be assembled and transported to the cell membrane, although free heavy chains can be found in the cytoplasm. Because the *BTK* gene is on the X chromosome, the disorder is seen in males.

Classically, this disease is characterized by the following:

- Absent or markedly decreased numbers of B cells in the circulation, with depressed serum levels of all classes of immunoglobulins. The numbers of pre-B cells in the bone marrow may be normal or reduced.
- Underdeveloped or rudimentary germinal centers in peripheral lymphoid tissues, including lymph nodes, Peyer patches, the appendix, and tonsils
- Absence of plasma cells throughout the body
- Normal T cell–mediated responses

XLA does not become apparent until the affected infant attains the age of approximately 6 months, when the transplacental supply of maternal antibodies is depleted. In most cases, recurrent bacterial infections such as acute and chronic pharyngitis, sinusitis, otitis media, bronchitis, and pneumonia suggest an underlying immune defect. The causal organisms typically are those bacterial pathogens that are cleared by antibody-mediated opsonization and phagocytosis (e.g., *Haemophilus influenzae, Streptococcus*

pneumoniae, and *Staphylococcus aureus*). Because antibodies are important for neutralizing viruses, patients with XLA also are susceptible to certain viral infections, especially those caused by enteroviruses. Similarly, *Giardia lamblia,* an intestinal protozoan usually neutralized by secreted IgA, cannot be efficiently cleared and causes persistent infections. Fortunately, replacement therapy with intravenous immunoglobulin (IVIG) from pooled human serum allows most patients to adequately combat bacterial infections. Patients with XLA clear some viral, fungal, and protozoal infections, because their T cell–mediated immunity is intact. For unclear reasons, autoimmune diseases (such as RA and dermatomyositis) occur in as many as 20% of patients with this disease.

Common Variable Immunodeficiency

Common variable immunodeficiency is an umbrella term for a heterogeneous group of disorders characterized by hypogammaglobulinemia, impaired antibody responses to infection (or vaccination), and increased susceptibility to infections. The clinical manifestations are superficially similar to those of XLA, but in common variable immunodeficiency, males and females are affected equally and the onset of symptoms is much later, in the second or third decade of life. The diagnosis usually is one of exclusion (after other causes of immune deficiency are ruled out). The estimated prevalence of the disease is about 1 in 50,000. Although most patients have normal numbers of mature B cells, plasma cells are absent, suggesting a block in antigen-stimulated B cell differentiation. The defective antibody production has been variably attributed to intrinsic B cell defects, deficient T cell help, or excessive T cell suppressive activity. Paradoxically, these patients are prone to develop a variety of autoimmune disorders (hemolytic anemia, pernicious anemia), as well as lymphoid tumors. The underlying mechanism of the antibody deficiency is variable (hence the name). Some patients with this disease have mutations in B cell receptors for certain growth factors, or in molecules involved in T cell–B cell interactions. However, the genetic basis of most cases of the disease is not known.

Isolated IgA Deficiency

The most common of all the primary immune deficiency diseases, IgA deficiency affects about 1 in 700 whites. As noted previously, IgA is the major immunoglobulin in mucosal secretions and is thus involved in defending the airways and the gastrointestinal tract. Although most people with this condition are asymptomatic, weakened mucosal defenses predispose patients to recurrent sinopulmonary infections and diarrhea. There is also a significant (but unexplained) association with autoimmune diseases. The pathogenesis of IgA deficiency seems to involve a block in the terminal differentiation of IgA-secreting B cells to plasma cells; IgM and IgG subclasses of antibodies are present in normal or even supranormal levels. The molecular basis for this defect is not understood.

Hyper-IgM Syndrome

In a normal immune response to protein antigen, IgM antibodies are produced first, followed by the sequential elaboration of IgG, IgA, and IgE antibodies. As discussed earlier in this chapter, the orderly appearance of different antibody types is called heavy-chain class (isotype) switching and is important for generating classes of antibody that can effectively activate complement and/or opsonize bacterial pathogens. The ability of IgM-producing B cells to turn on the transcription of genes that encode other immunoglobulin isotypes depends on certain cytokines, as well as on contact-mediated signals from CD4+ helper T cells. The contact-dependent signals are provided by interaction between CD40 molecules on B cells and CD40L (also known as CD154), expressed on activated helper T cells. *Patients with the hyper-IgM syndrome produce normal (or even supranormal) levels of IgM antibodies to antigens but lack the ability to produce the IgG, IgA, and IgE isotypes*; the underlying defect is an inability of T cells to induce B cell isotype switching. The most common genetic abnormality is mutation of the gene encoding CD40L. This gene is located on the X chromosome; consequently, *in approximately 70% of the cases, hyper-IgM syndrome is X-linked*. In the remaining patients, the mutations affect CD40 or other molecules involved in class switching, notably an enzyme called activation-induced deaminase. In addition to defective class switching, in those with CD40 or CD40L mutations there is also often a defect in the production of high-affinity antibodies, because the same mechanism is responsible for affinity maturation of the antibody response.

Although the disease is diagnosed and named because of the antibody abnormality, in patients with CD40 or CD40L mutations there is also a defect in cell-mediated immunity because the CD40–CD40L interaction is critical for helper T cell–mediated activation of macrophages, the central reaction of cell-mediated immunity. Male patients with the X-linked form of hyper-IgM syndrome present with recurrent pyogenic infections owing to low levels of opsonizing IgG antibodies. These patients also are susceptible to infections with a variety of intracellular pathogens that normally are combated by cell-mediated immunity, including *Pneumocystis jiroveci* (formerly called *Pneumocystis carinii*).

Thymic Hypoplasia: DiGeorge Syndrome

DiGeorge syndrome results from a congenital defect in thymic development with deficient T cell maturation. T cells are absent in the lymph nodes, spleen, and peripheral blood, and infants with this defect are extremely vulnerable to viral, fungal, and protozoal infections. Patients are also susceptible to infection with intracellular bacteria, because of defective T cell–mediated immunity. B cells and serum immunoglobulins are generally unaffected.

The disorder is a consequence of a developmental malformation affecting the third and fourth pharyngeal pouches, structures that give rise to the thymus, parathyroid glands, and portions of the face and aortic arch. Thus, in addition to the thymic and T cell defects, there may be parathyroid gland hypoplasia, resulting in hypocalcemic tetany, as well as additional midline developmental abnormalities. In 90% of cases of DiGeorge syndrome there is a deletion affecting chromosomal region 22q11, as discussed in Chapter 6. Transplantation of thymic tissue has successfully treated some affected infants. In patients with partial defects, immunity may improve spontaneously with age.

Severe Combined Immunodeficiency

Severe combined immunodeficiency (SCID) represents a constellation of genetically distinct syndromes with the common feature of defects in both humoral and cell-mediated immune responses. Affected infants are susceptible to severe recurrent infections by a wide array of pathogens, including bacteria, viruses, fungi, and protozoans, and opportunistic infections by *Candida, Pneumocystis,* CMV, and *Pseudomonas*. These pathogens cause serious (and occasionally lethal) disease. The prevalence of the disease is approximately 1 in 65,000 to 1 in 100,000, and the frequency is 20 to 30 times higher in some Native American populations.

Despite the common clinical features, the underlying defects in individual patients are quite diverse. Some forms of SCID are caused by a single defect affecting both T and B cells, and others may result from a primary T cell deficit with secondary impairment of humoral immunity. *Approximately half of the cases are X-linked; these are caused by mutations in the gene encoding the common γ chain* shared by the receptors for the cytokines IL-2, IL-4, IL-7, IL-9, and IL-15. Of these cytokines, IL-7 is the most important in this disease because it is the growth factor responsible for stimulating the survival and expansion of immature B and T cell precursors in the generative lymphoid organs. Another 40% to 50% of SCID cases are inherited in an autosomal recessive fashion, with approximately half of these caused by *mutations in adenosine deaminase (ADA)*, an enzyme involved in purine metabolism. ADA deficiency results in accumulation of adenosine and deoxyadenosine triphosphate metabolites, which inhibit DNA synthesis and are toxic to lymphocytes. The other autosomal recessive cases of SCID are attributed to defects in another purine metabolic pathway, primary failure of class II MHC expression, or mutations in genes encoding the recombinase responsible for the rearrangement of lymphocyte antigen-receptor genes.

In the two most common forms of SCID (cytokine receptor common γ chain mutation and ADA deficiency), the thymus is hypoplastic. Lymph nodes and lymphoid tissues (e.g., in the tonsils, gut, and appendix) are atrophic and lack germinal centers as well as paracortical T cells. Affected patients may have marked lymphopenia, with both T and B cell deficiency; others may have increased numbers of immature T cells and/or large numbers of B cells that are nonfunctional as a consequence of a lack of T cell help. Patients with SCID are currently treated by bone marrow transplantation. X-SCID is the first disease in which gene therapy has been used to successfully replace the mutated gene, but the approach is being reevaluated because some of the treated patients have developed T cell leukemias, presumably because the introduced gene was inserted close to a cellular oncogene.

Defects in Lymphocyte Activation

Rare patients with mutations in genes required for T cell activation have been identified that manifest with defective cell-mediated immunity or a phenotype resembling SCID. One of the interesting ones is a mutation in a transcription factor that is required for the T_H17 response. The resulting disease manifestations include fungal (and occasionally bacterial) skin infections and chronic mucocutaneous candidiasis. By contrast, mutations in genes involved in T_H1 responses result in susceptibility to infections with atypical mycobacteria. These observations emphasize the importance of T_H17 cells for defense against fungal infections and of T_H1 cells for combating intracellular bacterial infections. Mutations in genes encoding calcium channel proteins, and other components of T cell signaling, also have been described.

Immune Deficiency with Thrombocytopenia and Eczema: Wiskott-Aldrich Syndrome

Wiskott-Aldrich syndrome is an X-linked recessive disease *characterized by thrombocytopenia, eczema, and a marked vulnerability to recurrent infection*, ending in early death; the only treatment is bone marrow transplantation. This is a curious syndrome in that the clinical presentation and immunologic deficits are difficult to explain on the basis of the known underlying genetic defect. The thymus is initially normal, but there is progressive age-related depletion of T lymphocytes in the peripheral blood and lymph nodes, with concurrent loss of cellular immunity. Additionally, patients do not make effective antibody responses to polysaccharide antigens, and are therefore particularly susceptible to infections with encapsulated, pyogenic bacteria. Affected patients also are prone to the development of malignant lymphomas. The responsible gene encodes a protein (Wiskott-Aldrich syndrome protein) that links several membrane receptors to the cytoskeleton. Although the mechanism is not known, a defect in this protein could result in abnormal cellular morphology (including platelet shape changes) or defective cytoskeleton-dependent activation signals in lymphocytes and other leukocytes, with abnormal cell–cell adhesions and leukocyte migration.

Genetic Deficiencies of Components of Innate Immunity

Several genetic defects have been shown to affect molecules or cells that are important in the early innate immune response to microbes.

Complement Proteins

As discussed earlier in this chapter and in Chapter 2, complement components play important roles in inflammatory and immunologic responses. Consequently, hereditary deficiency of C3 (critical for both the classical and alternative pathways) results in an increased susceptibility to infection with pyogenic bacteria. Inherited deficiencies of C1q, C2, and C4 do not make affected persons susceptible to infections, but they do increase the risk of immune complex–mediated disease (e.g., SLE), possibly by impairing the clearance of apoptotic cells or of antigen–antibody complexes from the circulation. Deficiencies of the late components of the classical complement pathway (C5 to C9) result in recurrent infections by Neisseria (gonococci, meningococci) but not by other microbes because Neisseria have thin cell walls and are especially susceptible to lysis by the membrane attack complex, the product of the late steps of complement activation. Lack of the regulatory protein C1 inhibitor allows unfettered C1 activation, with the generation of downstream vasoactive complement mediators; the result is hereditary angioedema, characterized by recurrent episodes of localized edema affecting the skin and/or mucous membranes.

Phagocytes

Several congenital defects in phagocytes are known. These include defects in the phagocyte oxidase (NADPH oxidase) enzyme, the cause of chronic granulomatous disease, and defects in integrins and selectin ligands, causing the leukocyte adhesion deficiencies. These disorders were described in Chapter 2.

Other Genetic Disorders of Innate Immunity

Mutations in TLRs and their signaling pathways are quite rare, but study of the associated disorders has been informative. One of the surprises that has emerged from these diseases is that the immune deficiency typically is very restricted. For instance, patients with mutations affecting TLR3, which recognizes viral RNA, develop recurrent herpes simplex encephalitis, and those with mutations affecting MyD88, the signaling protein downstream of many TLRs, are susceptible to bacterial infections, especially severe pneumococcal lung disease, but neither suffers from multiple disseminated infections.

SUMMARY

Primary (Congenital) Immune Deficiency Diseases

- Caused by mutations in genes involved in lymphocyte maturation or function, or in innate immunity
- Some of the common disorders:
  - *XLA*: failure of B cell maturation, absence of antibodies; caused by mutations in BTK, which encodes a tyrosine kinase required for maturation signals from the pre-B cell and B cell receptors
  - *Common variable immunodeficiency*: defects in antibody production; cause unknown in most cases
  - *Selective IgA deficiency*: failure of IgA production; cause unknown
  - *X-SCID*: failure of T cell and B cell maturation; mutation in the common γ chain of a cytokine receptor, leading to failure of IL-7 signaling and defective lymphopoiesis
  - *Autosomal SCID*: failure of T cell development, secondary defect in antibody responses; approximately 50% of cases caused by mutation in the gene encoding ADA, leading to accumulation of toxic metabolites during lymphocyte maturation and proliferation
  - *X-linked hyper-IgM syndrome*: failure to produce isotype-switched high-affinity antibodies (IgG, IgA, IgE); mutation in gene encoding CD40L
- Clinical presentation: increased susceptibility to infections in early life

Secondary (Acquired) Immune Deficiencies

Immune deficiencies secondary to other diseases or therapies are much more common than the primary (inherited) disorders. Secondary immune deficiencies may be encountered in patients with malnutrition, infection, cancer, renal disease, or sarcoidosis. However, the most common cause of immune deficiency is therapy-induced suppression of the bone marrow or of lymphocyte function.

Discussed next is perhaps the most important secondary immune deficiency disease, AIDS, which has become one of the great scourges of humankind.

Acquired Immunodeficiency Syndrome (AIDS)

AIDS is a retroviral disease caused by the human immunodeficiency virus (HIV). It is characterized by infection and depletion of CD4+ T lymphocytes, and by profound immunosuppression leading to opportunistic infections, secondary neoplasms, and neurologic manifestations. Although AIDS was first described in the United States, it has now been reported in virtually every country in the world. At the end of 2009, more than 33 million people were living with HIV infection and AIDS, of which approximately 70% were in Africa and 20% in Asia; there were almost 2 million cases diagnosed and almost 2 million died of the disease that year, with a total of more than 22 million deaths since the epidemic was recognized in 1981. Although the largest number of infections is in Africa, the most rapid increases in HIV infection in the past decade have occurred in Southeast Asian countries, including Thailand, India, and Indonesia. The statistics are only slightly better in the Western world; for example, approximately 1 million U.S. citizens are infected (roughly 1 in 300). Moreover, more Americans (more than 500,000) have died of AIDS than died in both world wars combined. AIDS-related death rates continue to decline from their 1995 peak.

Because of the combined work of many scientists and clinicians, there has been an explosion of new knowledge about this modern plague. So rapid is the pace of research on the biology of HIV that any text covering the topic will probably be out of date by the time it goes to press. Nevertheless, presented next is a summary of the currently available information on HIV epidemiology, etiology, pathogenesis, and clinical features.

Epidemiology

Epidemiologic studies in the United States have identified five groups at risk for developing AIDS, and these are similar in other countries, except as noted in the following list. *Transmission of HIV occurs under conditions that facilitate the exchange of blood or body fluids that contain the virus or virus-infected cells.* Thus, the major routes of HIV infection are sexual contact, parenteral inoculation, and passage of the virus from infected mothers to their newborns. In about 10% of cases, the risk factors are unknown or not reported. The case distribution data cited are for the United States.

- Men who have sex with men constitute the largest group of infected persons, accounting for 48% of reported cases in the period 2001 to 2004 and 56% of infected men (approximately 4% of whom also inject drugs). However, transmission of AIDS in this category is declining, with less than 50% of new cases attributable to male homosexual contact.
- Heterosexual contacts of members of other high-risk groups constituted about 34% of infections from 2001 to 2004. In Africa and Asia, this is by far the largest group of patients with new infections, and a majority of new cases are in women infected by male partners.

- Intravenous drug abusers with no history of homosexual behavior compose the next largest group, representing about 17% of all patients.
- Recipients of blood and blood components (but not hemophiliacs) who received transfusions of HIV-infected whole blood or components (e.g., platelets, plasma) account for about 1% of patients.
- Hemophiliacs, especially those who received large amounts of factor VIII or IX concentrates before 1985, make up less than 1% of all cases.
- The epidemiology of HIV infection and AIDS is quite different in children (diagnosed when younger than 13 years of age). About 1% of all AIDS cases occur in this population, and the vast majority (about 90%) result from vertical transmission of virus from infected mother to the fetus or newborn.

Sexual Transmission. Sexual transmission is by far the major mode of infection worldwide, accounting for more than 75% of all cases of HIV transmission. Although most sexually transmitted cases in the United States are still due to male-with-male sexual contacts, the vast majority of sexually transmitted HIV infections globally are due to heterosexual activity. Even in the United States, the rate of increase of heterosexual transmission has outpaced transmission by other means; such spread accounts for the dramatic increase in HIV infection in female sex partners of male intravenous drug abusers.

The virus is present in semen, both extracellularly and within mononuclear inflammatory cells, and it enters the recipient's body through lacerations or abrasions in mucosa. Viral transmission to newborns can occur either by direct entry of virus or infected cells into blood vessels breached through traumatic injury or by uptake into mucosal DCs. Clearly, all forms of sexual transmission are aided and abetted by the concomitant presence of other sexually transmitted diseases that cause genital ulcerations, including syphilis, chancroid, and herpes simplex virus infections. Gonorrhea and chlamydial infection also are cofactors for HIV transmission, primarily by increasing the seminal fluid content of inflammatory cells (presumably carrying HIV). In addition to male-to-male and male-to-female transmission, HIV is present in the vaginal and cervical cells of infected women and can also be spread from females to males, albeit about eight times less efficiently.

Parenteral Transmission. Parenteral transmission of HIV is well documented in three different groups: intravenous drug abusers (the largest group), hemophiliacs receiving factor VIII or IX concentrates, and random recipients of blood transfusion. Among intravenous drug abusers, transmission occurs through shared needles, syringes, or other paraphernalia contaminated with HIV-containing blood.

Transmission of HIV by transfusion of blood or blood products such as lyophilized factor VIII concentrates has been virtually eliminated since 1985. Four public health measures are responsible: screening of donated blood and plasma for antibody to HIV, screening for HIV-associated p24 antigen (detectable before the development of antibodies), heat treatment of clotting factor concentrates, and screening of donors on the basis of history. With all of these measures, it is estimated currently that one in 1.5 million blood donations are HIV infected, and that 20 HIV-positive blood components derived from 11 infectious donations are released each year that could potentially infect recipients. With the advent of nucleic acid testing, this already small risk is expected to show further decline.

Mother-to-Infant Transmission. As noted earlier, mother-to-infant vertical transmission is the major cause of pediatric AIDS. Three routes are involved: in utero, by transplacental spread; intrapartum, during delivery; and by ingestion of HIV-contaminated breast milk. Of these, the transplacental and intrapartum routes account for most cases. Vertical transmission rates worldwide vary, ranging from 25% to 35%, with a 15% to 25% rate reported in the United States; higher rates of infection occur with high maternal viral load and/or the presence of chorioamnionitis, presumably by increasing placental accumulation of inflammatory cells.

Because of the dismal outcome for AIDS, the lay public is justifiably concerned about the spread of HIV infection outside recognized high-risk groups. Many of these anxieties can be laid to rest, because extensive studies indicate that HIV infection cannot be transmitted by casual personal contact in the home, workplace, or school, and no convincing evidence for spread by insect bites has been obtained. There is an extremely small but confirmed risk of transmission of HIV infection to health care workers. Seroconversion has been documented after accidental needlestick injury or exposure of nonintact skin to infected blood in laboratory accidents, with a rate of about 0.3% per accidental exposure. By comparison, the rate of seroconversion after accidental exposure to hepatitis B–infected blood is about 6% to 30%.

Etiology and Pathogenesis

AIDS is caused by HIV, a human retrovirus belonging to the lentivirus family (which also includes feline immunodeficiency virus, simian immunodeficiency virus, visna virus of sheep, and the equine infectious anemia virus). Two genetically different but antigenically related forms of HIV, called HIV-1 and HIV-2, have been isolated from patients with AIDS. HIV-1 is the more common type associated with AIDS in the United States, Europe, and Central Africa, whereas HIV-2 causes a similar disease principally in West Africa. Specific tests for HIV-2 are now available, and blood collected for transfusion also is routinely screened for HIV-2 seropositivity. The ensuing discussion relates primarily to HIV-1 and diseases caused by it, but it is generally applicable to HIV-2 as well.

Structure of HIV

Like most retroviruses, the HIV-1 virion is spherical and contains an electron-dense, cone-shaped core surrounded by a lipid envelope derived from the host cell membrane (Fig. 4–26). The virus core contains (1) major capsid protein p24, (2) nucleocapsid protein p7/p9, (3) two copies of genomic RNA, and (4) three viral enzymes—protease, reverse transcriptase, and integrase. The p24 protein is the most readily detected viral antigen and is therefore the target for the antibodies used to diagnose HIV infection in blood screening. The viral core is surrounded by a matrix protein called p17, lying beneath the virion envelope. The viral envelope itself is studded with two viral

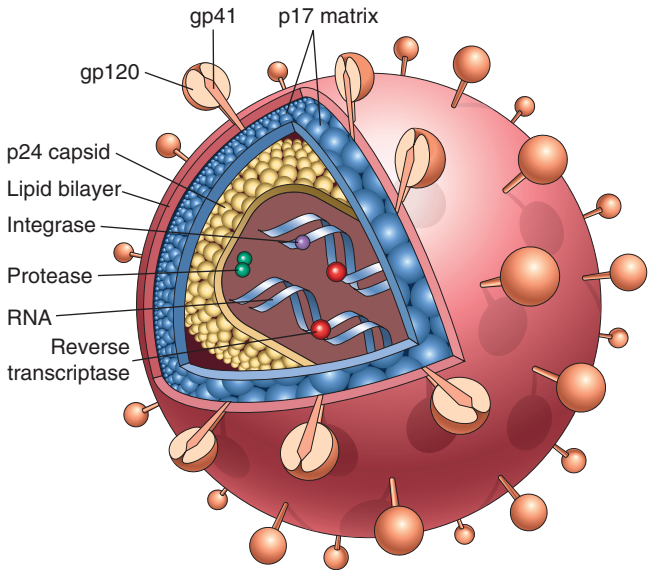


Figure 4–26 The structure of human immunodeficiency virus (HIV). The HIV-1 virion. The viral particle is covered by a lipid bilayer derived from the host cell and studded with viral glycoproteins gp41 and gp120.

glycoproteins (gp120 and gp41), critical for HIV infection of cells. The HIV-1 proviral genome contains the gag, pol, and env genes, which code for various viral proteins. The products of the gag and pol genes are translated initially into large precursor proteins that must be cleaved by the viral protease to yield the mature proteins. The highly effective anti-HIV-1 protease inhibitor drugs prevent viral assembly by inhibiting the formation of mature viral proteins.

In addition to these three standard retroviral genes, HIV contains several other genes (given three-letter names such as tat, rev, vif, nef, vpr, and vpu) that regulate the synthesis and assembly of infectious viral particles. The product of the tat (transactivator) gene, for example, is critical for virus replication, causing a 1000-fold increase in the transcription of viral genes. The nef protein activates intracellular kinase activity (affecting T cell activation, viral replication, and viral infectivity) and reduces surface expression of CD4 and MHC molecules on infected cells. The progression of HIV infection in vivo is dependent on nef; strains of simian immunodeficiency virus with mutated nef genes cause AIDS in monkeys at a markedly decreased rate, and humans infected with a nef-defective HIV-1 strain display low viral burden, with AIDS onset at a substantially slower pace than for nonmutant strains. The products of various regulatory genes are important for HIV pathogenicity, and therapeutic approaches are being developed to block their actions.

Nucleic acid sequencing of different viral isolates reveals considerable variability in many parts of the HIV genome. This high variability is due to the relatively low fidelity of the viral polymerase, with estimates of one mistake for each 10^5 replicated nucleotides. Most sequence variants cluster in parts of the genome encoding the envelope glycoproteins. Because the immune response against HIV-1 is targeted against its envelope, such extreme variability in antigen structure poses a formidable barrier to vaccine development.

On the basis of genomic analyses, HIV-1 can be divided into two groups, designated M (major) and O (outlier). Group M viruses, the more common form worldwide, are further divided into subtypes, or clades, designated A through J. The clades differ in their geographic distribution, with B being the most common form in Western Europe and the United States and E being the most common in Thailand. Beyond molecular homologies, the clades also show differences in modes of transmission. Thus, E clade is spread predominantly by heterosexual contact (male-to-female), presumably because of its ability to infect vaginal subepithelial DCs. By contrast, B clade virus grows poorly in DCs and may be transmitted by monocytes and lymphocytes.

Life Cycle of HIV

The two major targets of HIV infection are the immune system and the CNS. The life cycle of the virus is best understood in terms of its interactions with the immune system.

The entry of HIV into cells requires the CD4 molecule, which acts as a high-affinity receptor for the virus (Fig. 4–27). This requirement explains the tropism of the virus for CD4+ T cells and its ability to infect other CD4+ cells, particularly macrophages and DCs. However, binding to CD4 is not sufficient for infection; *the HIV envelope gp120 must also bind to other cell surface molecules (coreceptors) to facilitate cell entry*. Two cell surface chemokine receptors, CCR5 and CXCR4, serve this role. HIV envelope gp120 (noncovalently attached to transmembrane gp41) binds initially to CD4 molecules (Fig. 4–27). This binding leads to a conformational change that exposes a new recognition site on gp120 for the coreceptors CXCR4 (mostly on T cells) or CCR5 (mostly on macrophages). The gp41 then undergoes a conformational change that allows it to insert into the target membrane, and this process facilitates fusion of the virus with the cell. After fusion, the virus core containing the HIV genome enters the cytoplasm of the cell.

The coreceptors are critical components of the HIV infection process, and their discovery resolved some previously unexplained observations regarding HIV tropism. It had been known that HIV strains could be classified according to their relative ability to infect macrophages and/or CD4+ T cells. Macrophage-tropic (R5 virus) strains infect both monocytes/macrophages and freshly isolated peripheral blood T cells, whereas T cell–tropic (X4 virus) strains infect only activated T cell lines. This selectivity is now explained by selective coreceptor usage. R5 strains use CCR5 as their coreceptor, and, because CCR5 is expressed on both monocytes and T cells, these cells are susceptible to infection by R5 strains. Conversely, X4 strains bind to CXCR4, which is expressed on T cell lines (and not on monocytes/macrophages), so that only activated T cells are susceptible. Of interest, approximately 90% of HIV infections initially are transmitted by R5 strains. Over the course of infection, however, X4 viruses gradually accumulate; these are especially virulent and are responsible for T cell depletion in the final, rapid phase of disease progression. It is thought that during the course of HIV infection, R5 strains evolve into X4 strains, as a result of mutations in genes that encode gp120. Persons with defective CCR5 receptors (of U.S. whites, 20% are heterozygous and 1% are homozygous for the mutant CCR5) are relatively resistant to developing

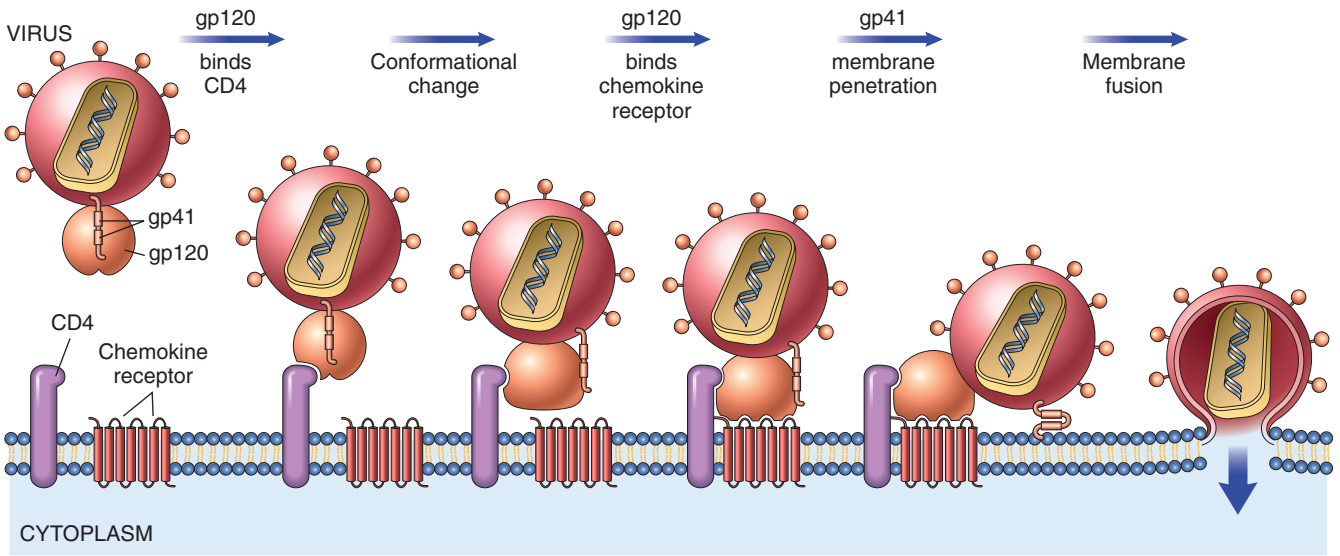


Figure 4–27 Molecular basis of entry of human immunodeficiency virus (HIV) into host cells. Interactions with CD4 and a chemokine receptor ("coreceptor").
(Adapted by permission from Macmillan Publishers Ltd, from Wain-Hobson S: HIV. One on one meets two. Nature 384:117, copyright 1996.)

AIDS, despite repeated HIV exposure in vivo. Because of the significance of HIV–coreceptor interaction in the pathogenesis of AIDS, preventing this interaction may be of significant therapeutic value.

Once internalized, the viral genome undergoes reverse transcription, leading to formation of complementary DNA (cDNA). In quiescent T cells, HIV proviral cDNA may remain in the cytoplasm in a linear episomal form. However, in dividing T cells, the cDNA enters the nucleus and becomes integrated into the host genome. After integration, the provirus may remain nontranscribed for months or years, and the infection becomes latent; alternatively, proviral DNA may be transcribed to form complete viral particles that bud from the cell membrane. Such productive infections, associated with extensive viral budding, lead to cell death. It is important to note that although HIV-1 can infect resting T cells, the initiation of proviral DNA transcription (and hence productive infection) occurs only when the infected cell is activated by exposure to antigens or cytokines. Thus, in a cruel twist, physiologic responses to infections and other stimuli promote the death of HIV-infected T cells.

Progression of HIV Infection

HIV disease begins with acute infection, which is only partly controlled by the host immune response, and advances to chronic progressive infection of peripheral lymphoid tissues (Fig. 4–28). The first cell types to be infected may be memory CD4+ T cells (which express CCR5) in mucosal lymphoid tissues. Because the mucosal tissues are the largest reservoir of T cells in the body and a major site of residence of memory T cells, the death of these cells results in considerable depletion of lymphocytes.

The transition from the acute phase to a chronic phase of infection is characterized by dissemination of the virus, viremia, and the development of host immune responses. DCs in epithelia at sites of virus entry capture the virus and then migrate into the lymph nodes. Once in lymphoid tissues, DCs may pass HIV on to CD4+ T cells through direct cell–cell contact. Within days after the first exposure to HIV, viral replication can be detected in the lymph nodes. This replication leads to viremia, during which high numbers of HIV particles are present in the patient's blood, accompanied by an acute HIV syndrome that includes a variety of nonspecific signs and symptoms typical of many viral diseases. The virus disseminates throughout the body and infects helper T cells, macrophages, and DCs in peripheral lymphoid tissues. As the infection spreads, the immune system mounts both humoral and cell-mediated immune responses directed at viral antigens. These immune responses partially control the infection and viral production, and such control is reflected by a drop in viremia to low but detectable levels by about 12 weeks after the primary exposure.

In the next, chronic phase of the disease, lymph nodes and the spleen are sites of continuous HIV replication and cell destruction (Fig. 4–28). During this period of the disease, the immune system remains competent at handling most infections with opportunistic microbes, and few or no clinical manifestations of the HIV infection are present. Therefore, this phase of HIV disease is called the *clinical latency period*. Although a majority of peripheral blood T cells do not harbor the virus, destruction of CD4+ T cells within lymphoid tissues steadily progresses during the latent period, and the number of circulating blood CD4+ T cells steadily declines. More than 90% of the body's approximately 10^{12} T cells are normally found in lymphoid tissues, and it is estimated that HIV destroys up to 1 to 2×10^9 CD4+ T cells every day. Early in the course of the disease, the body may continue to make new CD4+ T cells, so CD4+ T cells can be replaced almost as quickly as they are destroyed. At this stage, up to 10% of CD4+ T cells in lymphoid organs may be infected, but the number of circulating CD4+ T cells that are infected at any one time may be less than 0.1% of the total CD4+ T cells in a given case. Eventually, over a period of years, the continuous cycle of virus infection and T cell

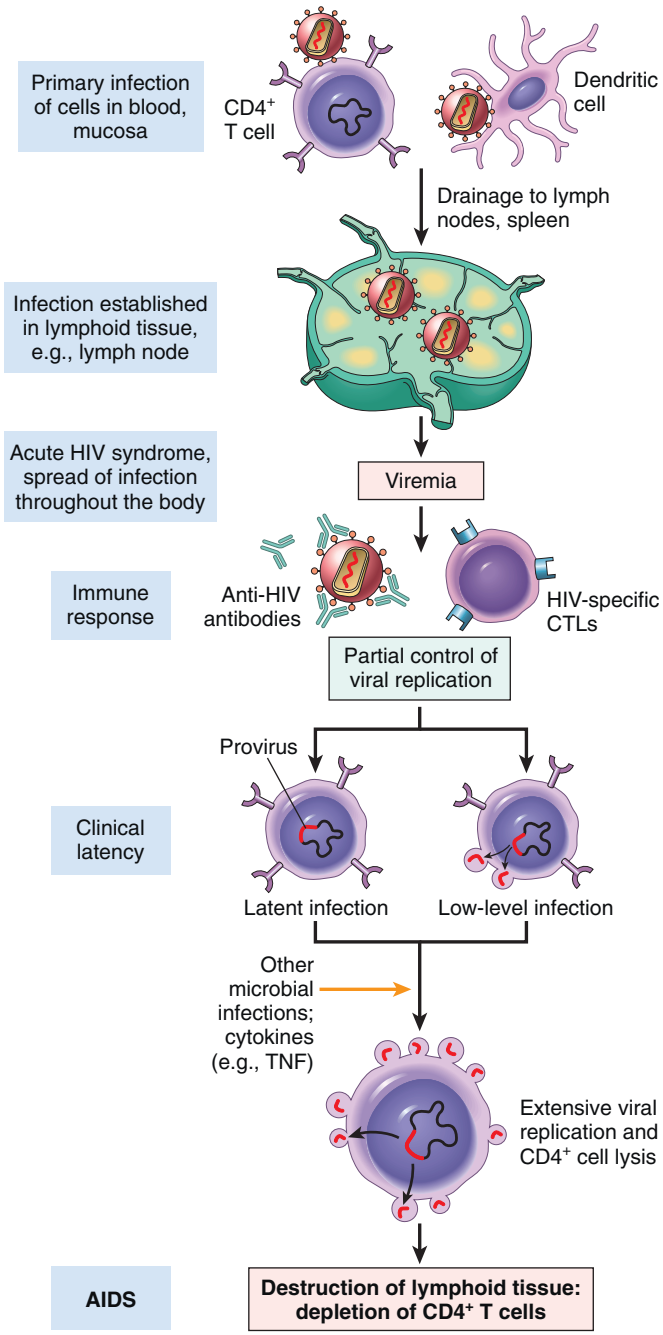


Figure 4–28 Pathogenesis of human immunodeficiency virus (HIV) infection. Initially, HIV infects T cells and macrophages directly or is carried to these cells by Langerhans cells. Viral replication in the regional lymph nodes leads to viremia and widespread seeding of lymphoid tissue. The viremia is controlled by the host immune response (not shown), and the patient then enters a phase of clinical latency. During this phase, viral replication in both T cells and macrophages continues unabated, but there is some immune containment of virus (not illustrated). There continues a gradual erosion of CD4+ cells by productive infection (or other mechanisms, not shown). Ultimately, CD4+ cell numbers decline and the patient develops clinical symptoms of full-blown AIDS. Macrophages are also parasitized by the virus early; they are not lysed by HIV and they transport the virus to tissues, particularly the brain.

death leads to a steady decline in the number of CD4+ T cells in the lymphoid tissues and the circulation.

In addition to T cell depletion, abnormalities have been described in many components of the immune system, summarized in Table 4–11. Discussed next are the major defects in immune cells during the course of HIV infection.

Mechanisms of T Cell Depletion in HIV Infection

The major mechanism of loss of CD4+ T cells is lytic HIV infection of the cells, and cell death during viral replication and production of virions (Fig. 4–29). Like other cytopathic viruses, HIV disrupts cellular functions sufficiently to cause death of infected cells. In addition to direct cell lysis, other mechanisms may cause T cell loss:

- Loss of immature precursors of CD4+ T cells, either by direct infection of thymic progenitor cells or by infection of accessory cells that secrete cytokines essential for CD4+ T cell maturation. The result is decreased production of mature CD4+ T cells.
- Chronic activation of uninfected cells by HIV antigens or by other concurrent infectious microbes may lead to apoptosis of the T cells. Because of this "activation-induced death" of uninfected cells, the numbers of T cells that die may be much greater than the number of HIV-infected cells.
- Infection of various cells in lymphoid tissues may disrupt the normal architecture, leading to impaired immune responses.
- Fusion of infected and uninfected cells causes formation of syncytia (giant cells). In tissue culture, the gp120 expressed on productively infected cells binds to CD4

Table 4–11 Major Abnormalities of Immune Function in AIDS

Lymphopenia
Predominantly caused by selective loss of the CD4+ helper T cell subset; reduced CD4+/CD8+ ratio
Decreased T Cell Function in vivo
Preferential loss of activated and memory T cells Decreased delayed-type hypersensitivity Susceptibility to opportunistic infections Susceptibility to neoplasms
Altered T Cell Function in vitro
Decreased proliferative response to mitogens, alloantigens, and soluble antigens Decreased cytotoxicity Decreased helper function for B cell antibody production Decreased interleukin-2 and interferon-γ production
Polyclonal B Cell Activation
Hypergammaglobulinemia and circulating immune complexes Inability to mount de novo antibody response to a new antigen Poor responses to normal signals for B cell activation in vitro
Altered Monocyte or Macrophage Functions
Decreased chemotaxis and phagocytosis Decreased HLA class II antigen expression Diminished capacity to present antigen to T cells Increased spontaneous secretion of interleukin-1, tumor necrosis factor, interleukin-6

HLA, human leukocyte antigen.

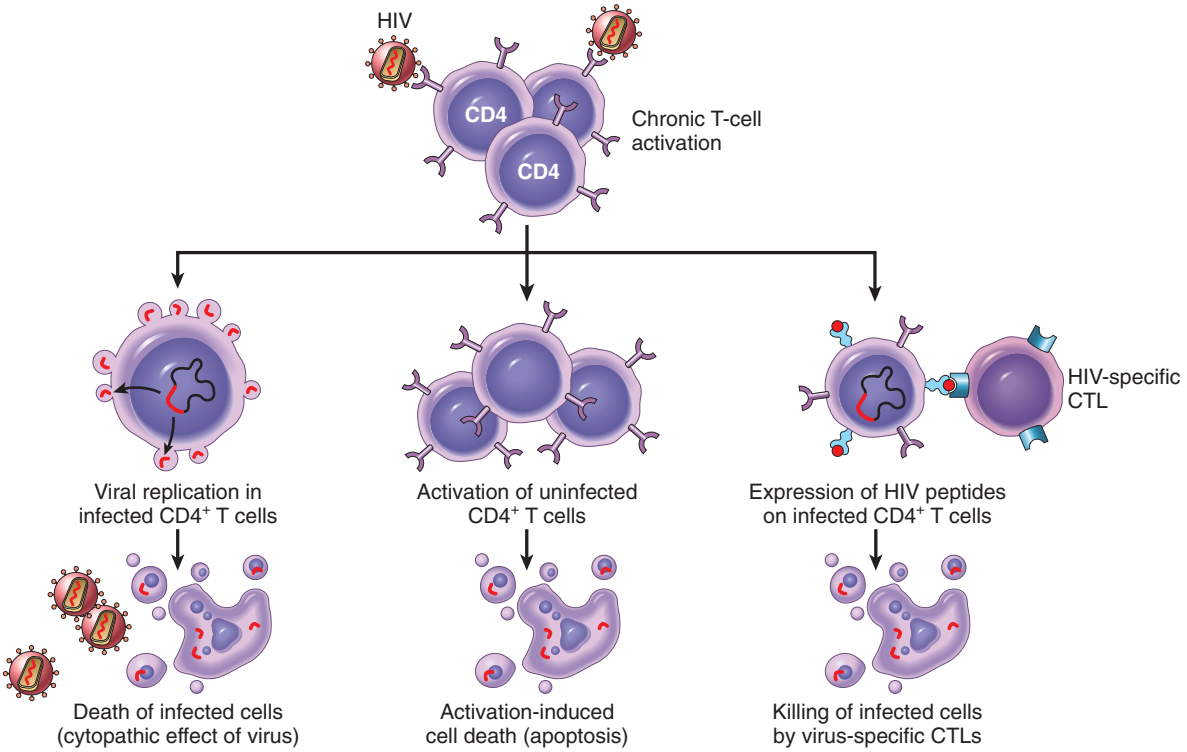


Figure 4–29 Mechanisms of CD4+ T cell loss in human immunodeficiency virus (HIV) infection. Some of the principal known and postulated mechanisms of T cell depletion after HIV infection are shown.

molecules on uninfected T cells, followed by cell fusion, ballooning, and death within a few hours. This property of syncytia formation is confined to the X4 strain of HIV.

- Uninfected CD4+ T cells may bind soluble gp120 to the CD4 molecule, leading to aberrant signaling and apoptosis.
- Infected CD4+ T cells may be killed by HIV-specific CD8+ CTLs.

The loss of CD4+ cells leads to an inversion of the CD4+/CD8+ ratio in the peripheral blood. Thus, while the normal CD4+/CD8+ ratio is 1 to 2, patients with AIDS have a ratio of 0.5 or less. Such inversion is a common finding in AIDS, but it also may occur in other viral infections and is therefore not diagnostic.

Although marked reduction in CD4+ T cells is a hallmark of AIDS and can account for much of the immune deficiency late in the course of HIV infection, there is also compelling evidence for qualitative defects in T cell function that can be detected even in asymptomatic HIV-infected persons. Such defects include reduced antigen-induced T cell proliferation, impaired T_H1 cytokine production, and abnormal intracellular signaling. There is also a selective loss of memory CD4+ T cells early in the course of the disease, possibly related to the abundance of these cells in mucosal tissues and higher level of CCR5 expression in this T cell subset.

Low-level chronic or latent infection of T cells (and macrophages) is an important feature of HIV infection. Although only rare CD4+ T cells express infectious virus early in the course of infection, up to 10% of lymph node T cells can be demonstrated to actually harbor the HIV genome. It is widely believed that integrated provirus, without virus production (latent infection), can persist within cells for months or years. Even with highly active antiretroviral therapy (which can eliminate most of the virus in the blood), latent virus lurks in lymph node CD4+ cells (as many as 0.05% of resting, long-lived CD4+ T cells are infected). Completion of the viral life cycle in latently infected cells requires cell activation. Thus, if latently infected CD4+ cells are activated by environmental antigens, an unfortunate consequence is increased HIV proviral DNA transcription. This increased transcription leads to virion production and, in the case of T cells, also results in cell lysis. In addition, TNF, IL-1, and IL-6 produced by activated macrophages during normal immune responses can also lead to increased HIV gene transcription (Fig. 4–28). Thus, it seems that HIV thrives when the host macrophages and T cells are physiologically activated (e.g., through intercurrent infection by other microbial agents). The life styles of most HIV-infected patients in the United States place them at increased risk for recurrent exposure to other sexually transmitted diseases; in Africa, socioeconomic conditions probably impose a higher burden of chronic microbial infections. It is easy to understand how patients with AIDS enter a vicious circle of T cell destruction; infections to which these patients are prone because of diminished helper T cell function lead to increased production of pro-inflammatory cytokines, which in turn stimulate more HIV production, followed by infection and loss of additional CD4+ T cells.

Monocytes/Macrophages in HIV Infection

In addition to infection of CD4+ T cells, infection of monocytes and macrophages is also important in the pathogenesis of HIV disease. Similar to T cells, most of the

HIV-infected macrophages are found in the tissues and not in peripheral blood. As many as 10% to 50% of macrophages in certain tissues, such as brain and lungs, may be infected. Several additional aspects of macrophage HIV infection warrant emphasis:

- Although cell division is required for integration and subsequent replication of most retroviruses, HIV-1 can infect and multiply in terminally differentiated nondividing macrophages, a property conferred by the HIV-1 vpr gene.
- Infected macrophages bud relatively small amounts of virus from the cell surface but contain large numbers of virus particles located in intracellular vesicles.
- In contrast with CD4+ T cells, macrophages are quite resistant to the cytopathic effects of HIV and can, therefore, harbor the virus for long periods.
- In more than 90% of cases, HIV infection is transmitted by R5 strains. The more virulent X4 strains that evolve later in the course of HIV infection are inefficient in transmitting HIV to monocytes. This suggests that the initial infection of macrophages (or DCs) is critical for HIV transmission.

Thus, in all likelihood, macrophages are the gatekeepers of HIV infection. Besides providing a portal for initial transmission, monocytes and macrophages are viral reservoirs and factories, whose output remains largely protected from host defenses. Circulating monocytes also provide a vehicle for HIV transport to various parts of the body, particularly the nervous system. In late stages of HIV infection, when the CD4+ T cell numbers are massively depleted, macrophages remain a major site of continued viral replication. Although the number of HIV-infected monocytes in the circulation is low, their functional deficits (e.g., impaired microbicidal activity, decreased chemotaxis, abnormal cytokine production, diminished antigen presentation capacity) have important bearing on host defenses.

DCs in HIV Infection

In addition to macrophages, two types of DCs also are important targets for the initiation and maintenance of HIV infection: mucosal and follicular DCs. As discussed earlier, mucosal DCs capture the virus and transport it to regional lymph nodes, where CD4+ T cells are infected. Follicular DCs in the germinal centers of lymph nodes are important reservoirs of HIV. Although some follicular DCs are infected by HIV, most virus particles are found on the surface of their dendritic processes, including those bound to Fc receptors through HIV–anti-HIV antibody complexes. The antibody-coated virions localized to follicular DCs retain the ability to infect CD4+ T cells. HIV infection of macrophages and DCs also may impair the functions of these cell populations, with secondary effects on T cell responsiveness.

B Cells and Other Lymphocytes in HIV Infection

Although much attention has been focused on T cells and macrophages, patients with AIDS also display profound abnormalities of B cell function. Paradoxically, these patients have hypergammaglobulinemia and circulating immune complexes as a result of polyclonal B cell activation. This may result from multiple factors, including infection with CMV or EBV, both of which are polyclonal B cell activators. The HIV gp41 itself can promote B cell growth and differentiation, and HIV-infected macrophages produce increased amounts of IL-6, which enhances B cell proliferation. Despite the presence of spontaneously activated B cells, patients with AIDS are unable to mount antibody responses to newly encountered antigens. Not only is this attributable to deficient T cell help, but antibody responses against T cell–independent antigens are also suppressed, suggesting additional B cell defects. Impaired humoral immunity renders these patients susceptible to encapsulated bacteria (e.g., *S. pneumoniae, H. influenzae*) that require antibodies for effective opsonization and clearance.

CD4+ T cells play a pivotal role in regulating the immune response: they produce a plethora of cytokines, chemotactic factors, and hematopoietic growth factors (e.g., granulocyte-macrophage colony-stimulating factor). Therefore, loss of this "master cell" has ripple effects on virtually every other cell of the immune system, as summarized in Table 4–11.

Pathogenesis of CNS Involvement

The pathogenesis of the neurologic manifestations in AIDS deserves special mention because, in addition to the lymphoid system, the nervous system is a major target of HIV infection. Macrophages and cells belonging to the monocyte–macrophage lineage (microglial cells) are the predominant cell types in the brain that are infected with HIV. The virus is most likely carried into the brain by infected monocytes (thus, brain HIV isolates are almost exclusively of the R5 type). The mechanism of HIV-induced damage of the brain, however, remains obscure. Because neurons are not infected by HIV, and the extent of neuropathologic changes is often less than might be expected from the severity of neurologic symptoms, most experts believe that the neurologic deficit is caused indirectly by viral products and soluble factors (e.g., cytokines such as TNF) produced by macrophages and microglial cells. In addition, injury from nitric oxide induced in neuronal cells by gp41 and direct damage of neurons by soluble HIV gp120 have been postulated.

SUMMARY

Human Immunodeficiency Virus Life Cycle and the Pathogenesis of AIDS

- Virus entry into cells: requires CD4 and co-receptors, which are receptors for chemokines; involves binding of viral gp120 and fusion with the cell mediated by viral gp41 protein; main cellular targets: CD4+ helper T cells, macrophages, DCs
- Viral replication: integration of provirus genome into host cell DNA; triggering of viral gene expression by stimuli that activate infected cells (e.g., infectious microbes, cytokines produced during normal immune responses)
- Progression of infection: acute infection of mucosal T cells and DCs; viremia with dissemination of virus; latent infection of cells in lymphoid tissue; continuing viral replication and progressive loss of CD4+ T cells

- Mechanisms of immune deficiency:
  - Loss of CD4+ T cells: T cell death during viral replication and budding (similar to other cytopathic infections); apoptosis occurring as a result of chronic stimulation; decreased thymic output; functional defects
  - Defective macrophage and DC functions
  - Destruction of architecture of lymphoid tissues (late)

Natural History and Clinical Course

The clinical course of HIV infection can best be understood in terms of an interplay between HIV and the immune system. Three phases reflecting the dynamics of virus–host interaction can be recognized: (1) an early acute phase, (2) a middle chronic phase, and (3) a final crisis phase (Fig. 4–30).

- The *acute phase* represents the initial response of an immunocompetent adult to HIV infection. Clinically, this phase typically manifests as a self-limited illness that develops in 50% to 70% of affected persons 3 to 6 weeks after infection; it is characterized by nonspecific symptoms including sore throat, myalgia, fever, rash, and sometimes aseptic meningitis. This phase is also characterized by high levels of virus production, viremia, and widespread seeding of the peripheral lymphoid tissues, typically with a modest reduction in CD4+ T cells. Soon, however, a virus-specific immune response develops, evidenced by seroconversion (usually within 3 to 17 weeks of exposure) and by the development of virus-specific CD8+ CTLs. As viremia abates, CD4+ T cells return to nearly normal numbers. However, the reduction in plasma virus does not signal the end of viral replication, which continues within CD4+ T cells and macrophages in the tissues (particularly lymphoid organs).
- The middle, *chronic phase* represents a stage of relative containment of the virus. The immune system is largely intact at this point, but there is continued HIV replication that may last for several years. Patients either are asymptomatic or develop persistent lymphadenopathy, and "minor" opportunistic infections such as thrush (*Candida*) or herpes zoster. During this phase, viral replication in the lymphoid tissues continues unabated; thus, there is no true microbiologic latency in HIV infection. The extensive viral turnover is associated with continued loss of CD4+ cells, but a large proportion of the CD4+ cells is replenished and the decline of CD4+ cells in the peripheral blood is modest. After an extended and variable period, the number of CD4+ cells begins to decline, the proportion of the surviving CD4+ cells infected with HIV increases, and host defenses begin to wane. Persistent lymphadenopathy with significant constitutional symptoms (fever, rash, fatigue) reflects the onset of immune system decompensation, escalation of viral replication, and the onset of the "crisis" phase.
- The final, *crisis phase* is characterized by a catastrophic breakdown of host defenses, a marked increase in viremia, and clinical disease. Typically, patients present with fever of more than 1 month's duration, fatigue, weight loss, and diarrhea; the CD4+ cell count is reduced below 500 cells/μL. After a variable interval, serious opportunistic infections, secondary neoplasms, and/or neurologic manifestations (so-called AIDS-defining conditions) emerge, and the patient is said to have full-blown AIDS. Even if the usual AIDS-defining conditions are not present, Centers for Disease Control and

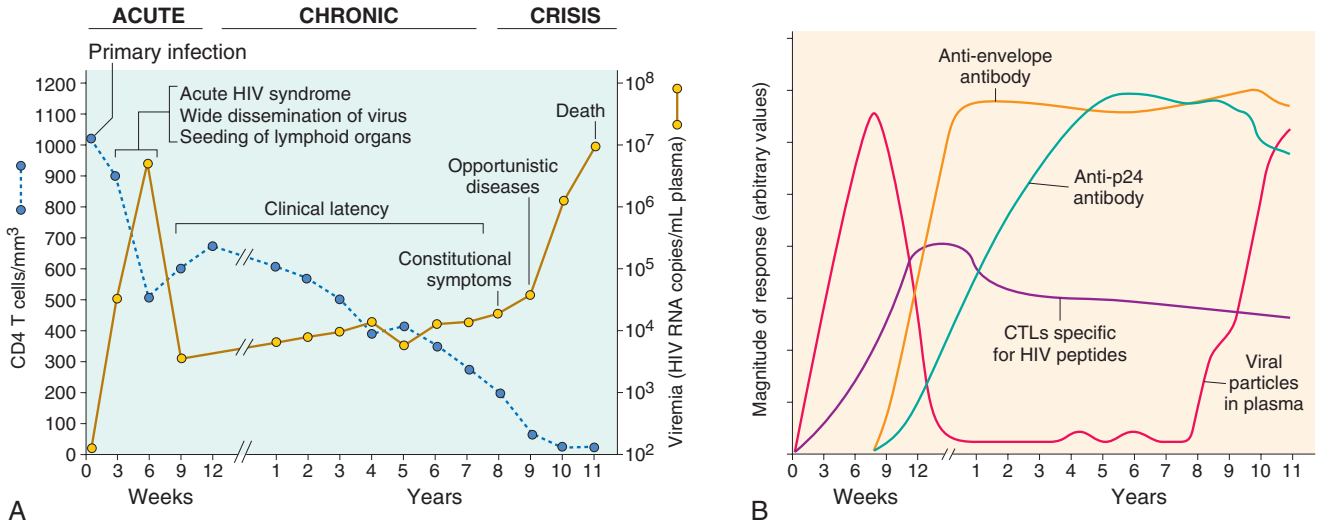


Figure 4–30 Clinical and immune response to human immunodeficiency virus (HIV) infection. **A,** Clinical course. The early period after primary infection is characterized by dissemination of virus, development of an immune response to HIV, and often an acute viral syndrome. During the period of clinical latency, viral replication continues, and the CD4+ T cell count gradually decreases until it reaches a critical level below which there is a substantial risk of AIDS-associated diseases. **B,** Immune response to HIV infection. A cytotoxic T lymphocyte (CTL) response to HIV is detectable by 2 to 3 weeks after the initial infection and peaks by 9 to 12 weeks. Marked expansion of virus-specific CD8+ T cell clones occurs during this time, and up to 10% of a patient's CTLs may be HIV-specific at 12 weeks. The humoral immune response to HIV peaks at about 12 weeks.

(A, Redrawn from Fauci AS, Lane HC: Human immunodeficiency virus disease: AIDS and related conditions. In Fauci AS, Braunwald E, Isselbacher KJ, et al [eds]: Harrison's Principles of Internal Medicine, 14th ed. New York, McGraw-Hill, 1997, p 1791.)

Prevention (CDC) guidelines define any HIV-infected person with CD4+ counts of 200 cells/μL or less as having AIDS.

In the absence of treatment, most patients with HIV infection develop AIDS after a chronic phase lasting 7 to 10 years. Exceptions to this time frame are seen in so-called rapid progressors and long-term nonprogressors. In rapid progressors, the middle, chronic phase is telescoped to 2 to 3 years after primary infection. Nonprogressors (less than 5% of infected persons) are defined as HIV-infected patients who remain asymptomatic for 10 years or more, with stable CD4+ counts and low levels of plasma viremia; notably, AIDS eventually develops in a majority of these patients, albeit after a much-prolonged clinical latency. Despite much study, the reason for nonprogression is not known.

Because the loss of immune containment is associated with declining numbers of CD4+ T cells, the CDC classification of HIV infection stratifies patients into three categories on the basis of CD4+ T cell counts: more than 500 cells/μL, between 200 and 500 cells/μL, and less than 200 cells/μL. Patients in the first group are generally asymptomatic; counts below 500 cells/μL are associated with early symptoms, and a decline of CD4+ T cell levels below 200 cells/μL is associated with severe immunosuppression. For clinical management, CD4+ cell counts are an important adjunct to HIV viral load measurements. The significance of these two measurements, however, is slightly different: Whereas CD4+ cell counts indicate the status of the patient's disease at the time of measurement, the viral load provides information about the direction in which the disease is progressing.

Although this summary of the clinical course is true for untreated or refractory cases, recently developed antiretroviral therapy has changed the course of the disease and greatly reduced the incidence of severe opportunistic infections (such as *Pneumocystis* pneumonia) and tumors (such as Kaposi sarcoma). The available therapy does not eliminate all of the virus, however, and the disease can recur if treatment is stopped. Also not known is whether drug-resistant viral strains will become widespread.

Clinical Features

The clinical manifestations of HIV infection range from a mild acute illness to severe disease. Because the salient clinical features of the acute, early and chronic, middle phases of HIV infection were described earlier, only the clinical manifestations of the terminal phase, full-blown AIDS, are summarized here.

In the United States the typical adult patient with AIDS presents with fever, weight loss, diarrhea, generalized lymphadenopathy, multiple opportunistic infections, neurologic disease, and (in many cases) secondary neoplasms. The infections and neoplasms listed in Table 4–12 are included in the surveillance definition of AIDS.

Opportunistic Infections. Opportunistic infections have accounted for approximately 80% of deaths among patients with AIDS. Their spectrum is constantly changing, and their incidence is decreasing markedly as a result of more effective antiretroviral therapy. A brief summary of selected opportunistic infections is provided here.

Pneumonia caused by the opportunistic fungus *P. jiroveci* (representing reactivation of a previous latent infection) is the presenting feature in many cases, although its incidence is declining as a result of effective prophylactic regimens. The risk of developing this infection is extremely high in individuals with fewer than 200 CD4+ T cells/μL. Many patients present with an opportunistic infection other than *P. jiroveci* pneumonia (Table 4–12). Among the most common are recurrent mucosal candidiasis, disseminated CMV infection (particularly enteritis and retinitis), severe ulcerating oral and perianal herpes simplex, and disseminated infection with M. tuberculosis and atypical mycobacteria (*Mycobacterium avium-intracellulare*). The AIDS epidemic has caused a resurgence of active tuberculosis in the United States. Although in most cases it represents reactivation, the frequency of new infections is also increasing. Whereas M. tuberculosis manifests itself early in the course of AIDS, infections with atypical mycobacteria are seen late in the course of HIV disease, usually occurring in patients with fewer than 100 CD4+ cells/μL. Toxoplasmosis is the most common secondary infection of the CNS. Cryptococcal meningitis also is quite frequent. Persistent diarrhea, which is common in patients with AIDS, often is caused by *Cryptosporidium* or *Isospora belli* infections, but bacterial pathogens such as *Salmonella* and *Shigella* also may be involved. Because of depressed humoral immunity, patients with AIDS are susceptible to infections with *S. pneumoniae* and *H. influenzae.*

Neoplasms. Patients with AIDS have a high incidence of certain tumors, particularly Kaposi sarcoma, non-Hodgkin lymphomas, and cervical cancer in women. The common feature of all of these diverse neoplasms is that the tumor cells in each are typically infected by an oncogenic virus. The basis of the increased risk of virus-associated

Table 4–12 AIDS-Defining Opportunistic Infections and Neoplasms Found in Patients with Human Immunodeficiency Virus (HIV) infection

Infections
Protozoal and Helminthic Infections
Cryptosporidiosis or isosporidiosis (enteritis)
Pneumocystosis (pneumonia or disseminated infection)
Toxoplasmosis (pneumonia or CNS infection)
Fungal Infections
Candidiasis (esophageal, tracheal, or pulmonary)
Cryptococcosis (CNS infection)
Coccidioidomycosis (disseminated)
Histoplasmosis (disseminated)
Bacterial Infections
Mycobacteriosis ("atypical," e.g., *Mycobacterium avium-intracellulare*, disseminated or extrapulmonary; *Mycobacterium tuberculosis*, pulmonary or extrapulmonary)
Nocardiosis (pneumonia, meningitis, disseminated)
Salmonella infections, disseminated
Viral Infections
Cytomegalovirus (pulmonary, intestinal, retinitis, or CNS infections)
Herpes simplex virus (localized or disseminated infection)
Varicella-zoster virus (localized or disseminated infection)
Progressive multifocal leukoencephalopathy
Neoplasms
Kaposi sarcoma
Primary lymphoma of brain
Invasive cancer of uterine cervix

CNS, central nervous system.

malignancy is multifactorial, but defective T cell immunity is believed to be the predominant contributor.

Kaposi sarcoma, a vascular tumor that is otherwise rare in the United States (Chapter 9), was once the most common neoplasm in AIDS patients but its incidence has decreased significantly with anti-retroviral therapy. The tumor is far more common among homosexual or bisexual males than in intravenous drug abusers or patients belonging to other risk groups. The lesions can arise early, before the immune system is compromised, or in advanced stages of HIV infection. Unlike the lesions in sporadic cases of Kaposi sarcoma, those that occur in patients with AIDS are multicentric and tend to be more aggressive; they can affect the skin, mucous membranes, gastrointestinal tract, lymph nodes, and lungs. The lesions contain spindle cells that share features with endothelial cells and smooth muscle cells and are believed to be lymphatic endothelial cells or mesenchymal cells that can form vascular channels. In different patients, the lesions are monoclonal or oligoclonal or even polyclonal, an unusual feature shared by other proliferations driven by oncogenic viruses, such as certain EBV-related B cell proliferations.

Kaposi sarcoma is caused by a herpesvirus called Kaposi sarcoma herpesvirus (KSHV), or human herpesvirus-8 (HHV-8). The mechanisms by which the virus causes the vascular proliferation are uncertain. One hypothesis is that KSHV infects lymphatic endothelial or other cells, and in concert with cytokines produced by HIV-infected immune cells, stimulates proliferation of the endothelial cells. The KSHV genome contains homologues of several human oncogenes and cytokines that may contribute to the growth and survival of the proliferating vessels.

B cell non-Hodgkin lymphomas constitute the second most common type of AIDS-associated tumors. These tumors are highly aggressive, occur most frequently in severely immunosuppressed patients, and involve many extranodal sites. The brain is the most common extranodal site in late-stage HIV infection, and hence primary lymphoma of the brain is considered an AIDS-defining condition. Close to 100% of these brain lymphomas are EBV-related. In comparison only 30% to 40% of lymphomas occuring earlier in the course of HIV infection are EBV-related, emphasizing the contribution of other factors, such as chronic B cell hyperstimulation, to lymphoma risk in HIV-infected individuals. Another, less common AIDS-related lymphoma is primary effusion lymphoma, which grows exclusively in body cavities, manifesting as pleural, peritoneal, or pericardial effusions. This rare tumor is always associated with KSHV, and in many cases the tumor cells are co-infected with both KSHV and EBV.

The incidence of *cervical carcinoma* also is increased in patients with AIDS. This correlation is attributable to the high prevalence of human papillomavirus infection among patients with AIDS, whose immune systems are compromised. This virus is believed to be intimately associated with squamous cell carcinoma of the cervix and its precursor lesions, cervical dysplasia and carcinoma in situ (Chapter 18). Hence, gynecologic examination should be part of the routine evaluation in HIV-infected women.

In general, the incidence of the classical "AIDS-defining cancers"—Kaposi sarcoma, EBV-associated tumors, and cervical cancer—has decreased significantly with the use of antiretroviral therapy, but the relative incidence of other tumors considered "non–AIDS-defining cancers" is actually increasing. This latter group includes liver cancer, anal cancer, and Hodgkin lymphoma, all of which are types of tumors associated with various viral infections.

CNS Involvement. Involvement of the CNS is a common and important manifestation of AIDS. At autopsy, 90% of patients are found to have some form of neurologic involvement, and 40% to 60% have clinically evident neurologic dysfunction. Significantly, in some patients neurologic manifestations may be the sole or earliest presenting feature of HIV infection. In addition to opportunistic infections and neoplasms, several virally determined neuropathologic changes occur. These include an aseptic meningitis occurring at the time of seroconversion, vacuolar myelopathy, peripheral neuropathies, and (most commonly) a progressive encephalopathy clinically designated the AIDS dementia complex (Chapter 22).

MORPHOLOGY

The anatomic changes in the tissues (with the exception of lesions in the brain) are neither specific nor diagnostic. In general, the pathologic features of AIDS are those of widespread opportunistic infections, Kaposi sarcoma, and lymphoma. Most of these lesions are discussed elsewhere, because they also occur in patients who do not have HIV infection. To appreciate the distinctive nature of lesions in the CNS, they are discussed in the context of other disorders affecting the brain (Chapter 22). Here the focus is on changes in the lymphoid organs.

Biopsy specimens from enlarged lymph nodes in the early stages of HIV infection reveal a **marked follicular hyperplasia** (Chapter 11). The medulla contains abundant **plasma cells**. These changes, affecting primarily the B cell areas of the node, are the morphologic counterparts of the polyclonal B cell activation and hypergammaglobulinemia seen in AIDS patients. In addition to changes in the follicles, the sinuses show increased cellularity, due primarily to increased numbers of macrophages but also contributed to by B cell lymphoblasts and plasma cells. HIV particles can be demonstrated within the germinal centers, concentrated on the villous processes of the follicular DCs. Viral DNA also can be detected in macrophages and CD4+ T cells.

With disease progression, the frenzy of B cell proliferation gives way to a pattern of severe follicular involution and generalized lymphocyte depletion. The organized network of follicular DCs is disrupted, and the follicles may even become hyalinized. These "burnt-out" lymph nodes are atrophic and small and may harbor numerous opportunistic pathogens. Because of profound immunosuppression, the inflammatory response to infections both in the lymph nodes and at extranodal sites may be sparse or atypical. For example, with severe immunosuppression, mycobacteria do not evoke granuloma formation, because CD4+ T cells are lacking. In the empty-looking lymph nodes and in other organs, the presence of infectious agents may not be readily apparent without the application of special stains. As might be expected, lymphoid depletion is not confined to the nodes; in the later stages of AIDS, the spleen and thymus also appear to be "wastelands."

Non-Hodgkin lymphomas, often involving extranodal sites such as the brain are primarily aggressive B cell neoplasms (Chapter 11).

Since the emergence of AIDS in 1981, the concerted efforts of epidemiologists, immunologists, and molecular biologists have resulted in spectacular advances in our understanding of this disorder. Despite all this progress, however, the prognosis of patients with AIDS remains guarded. Although the mortality rate has declined as a result of the use of potent combinations of antiretroviral drugs, all treated patients still carry viral DNA in their lymphoid tissues. Can there be a cure with persistent virus? Despite the considerable effort that has been mounted to develop a vaccine, many hurdles remain to be crossed before vaccine-based prophylaxis or treatment becomes a reality. Molecular analyses have revealed an alarming degree of variation in viral isolates from different patients, rendering vaccine development even more difficult. A further complication to this task is that the nature of the protective immune response is not yet fully understood. Consequently, at present, prevention and effective public health measures, combined with antiretroviral therapy, are the mainstays in the fight against AIDS.

AMYLOIDOSIS

Amyloidosis is a condition associated with a number of inherited and inflammatory disorders in which extracellular deposits of fibrillar proteins are responsible for tissue damage and functional compromise. These abnormal fibrils are produced by the aggregation of misfolded proteins (which are soluble in their normal folded configuration) or protein fragments. The fibrillar deposits bind a wide variety of proteoglycans and glycosaminoglycans, including heparan sulfate and dermatan sulfate, and plasma proteins, notably serum amyloid P component (SAP). The presence of abundant charged sugar groups in these adsorbed proteins gives the deposits staining characteristics that were thought to resemble starch (amylose). Therefore, the deposits were called "amyloid," a name that is firmly entrenched despite the realization that the deposits are unrelated to starch.

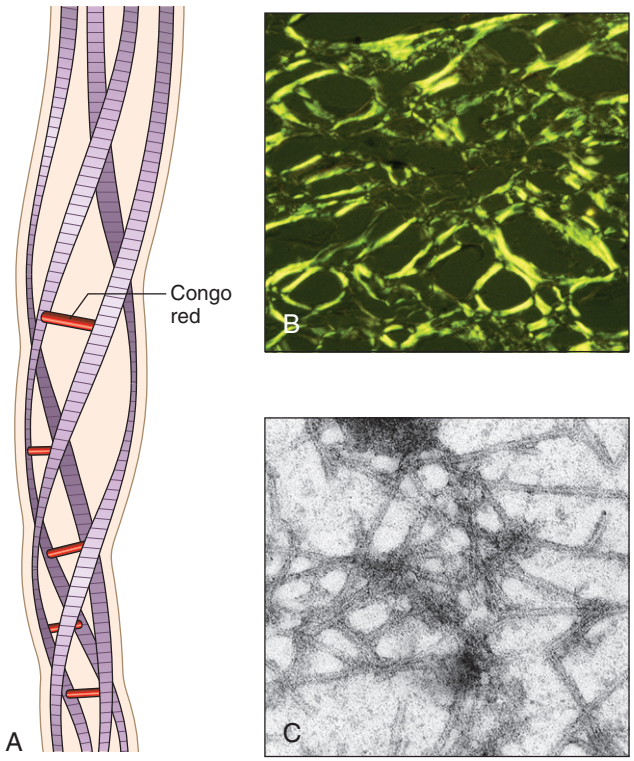


Figure 4–31 Structure of amyloid. **A,** Schematic diagram of an amyloid fiber showing fibrils (four are shown; as many as six may be present) wound around one another with regularly spaced binding of the Congo red dye. **B,** Congo red staining shows an apple-green birefringence under polarized light, a diagnostic feature of amyloid. **C,** Electron micrograph of 7.4- to 10-nm amyloid fibrils.

(Reproduced from Merlini G, Bellotti V: Molecular mechanisms of amyloidosis. N Engl J Med 349:583–596, 2003. Copyright 2003 Massachusetts Medical Society. All rights reserved.)

PATHOGENESIS OF AMYLOID DEPOSITION

Amyloidosis is fundamentally a disorder of protein misfolding. Amyloid is not a structurally homogeneous protein, although it always has the same morphologic appearance. In fact, more than 20 (at last count, 23) different proteins can aggregate to form fibrils with the appearance of amyloid. Regardless of their derivation, all amyloid deposits are composed of nonbranching fibrils, 7.5 to 10 nm in diameter, each formed of β-sheet polypeptide chains that are wound together (Fig. 4–31). The dye Congo red binds to these fibrils and produces a red–green dichroism (birefringence), which is commonly used to identify amyloid deposits in tissues.

Amyloidosis results from abnormal folding of proteins, which are deposited as fibrils in extracellular tissues and disrupt normal function. Normally, misfolded proteins are degraded intracellularly in proteasomes, or extracellularly by macrophages. It appears that in amyloidosis, these quality control mechanisms fail, allowing the misfolded protein to accumulate outside cells. Misfolded proteins often are unstable and self-associate, ultimately leading to the formation of oligomers and fibrils that are deposited in tissues. The diverse conditions that are associated with amyloidosis all are likely to result in excessive production of proteins that are prone to misfolding (Fig. 4–32). The proteins that form amyloid fall into two general categories: (1) normal proteins that have an inherent tendency to fold improperly, associate to form fibrils, and do so when they are produced in increased amounts and (2) mutant proteins that are prone to misfolding and subsequent aggregation. Of the many biochemically distinct forms of amyloid proteins that have been identified, three are most common:

- The **AL (amyloid light chain) protein** is produced by plasma cells and is made up of complete immunoglobulin light chains, the amino-terminal fragments of light chains, or both. For unknown reasons, only a few types of immunoglobulin light chains are prone to forming aggregates. As expected, the deposition of amyloid fibril protein of the AL type is associated with some form of monoclonal B cell proliferation. Defective degradation has also been invoked as the basis for fibril formation, and perhaps particular light chains are resistant to complete proteolysis. However, there are no sequence motifs peculiar to the immunoglobulin light chains found in amyloid deposits.
- The **AA (amyloid-associated) fibril** is a unique nonimmunoglobulin protein derived from a larger (12-kDa)

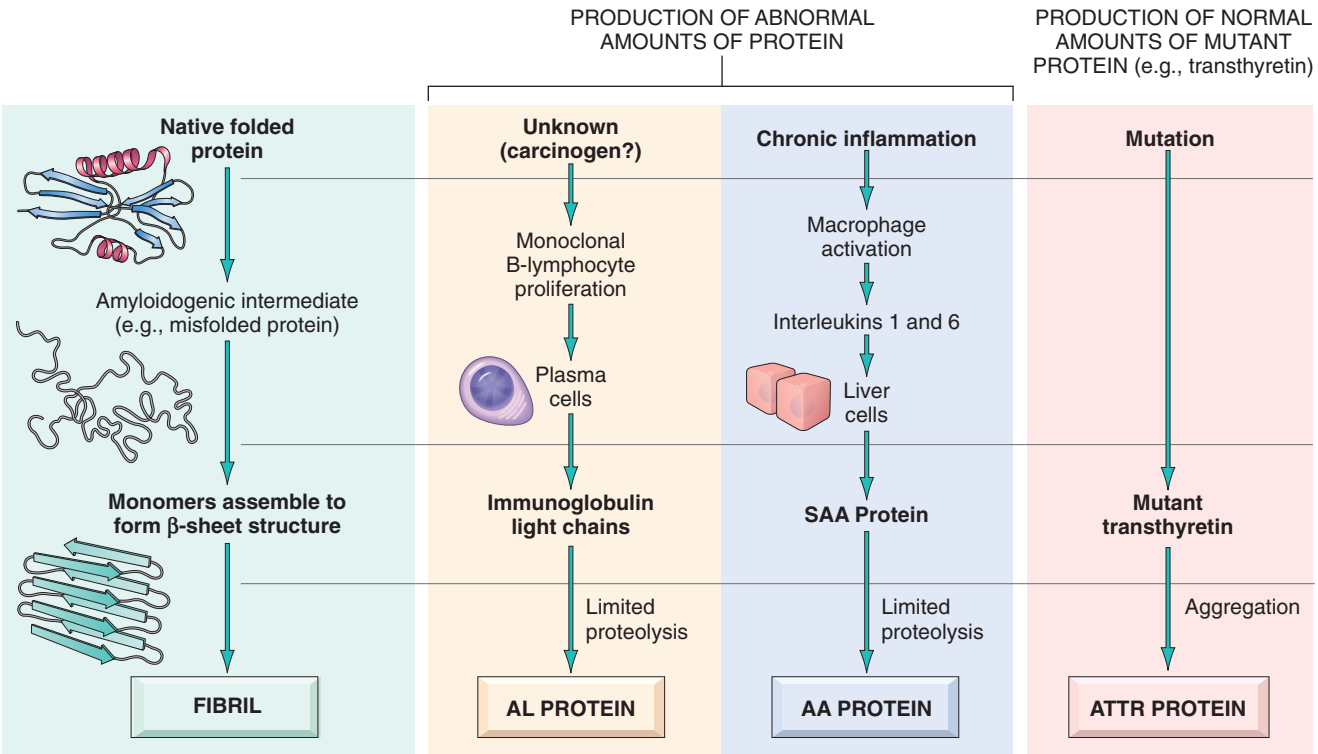


Figure 4–32 Pathogenesis of amyloidosis. The proposed mechanisms underlying deposition of the major forms of amyloid fibrils.

serum precursor called SAA (serum amyloid-associated) protein that is synthesized in the liver. SAA is synthesized by liver cells under the influence of cytokines such as IL-6 and IL-1 that are produced during inflammation; thus, long-standing inflammation leads to elevated SAA levels, and ultimately the AA form of amyloid deposits. However, increased production of SAA by itself is not sufficient for the deposition of amyloid. Elevation of serum SAA levels is common to inflammatory states but in most instances does not lead to amyloidosis. There are two possible explanations for this. According to one view, SAA normally is degraded to soluble end products by the action of monocyte-derived enzymes. Conceivably, people who develop amyloidosis have an enzyme defect that results in incomplete breakdown of SAA, thus generating insoluble AA molecules. Alternatively, a genetically determined structural abnormality in the SAA molecule itself renders it resistant to degradation by macrophages.

- **Aβ amyloid** is found in the cerebral lesions of Alzheimer disease. Aβ is a 4-kDa peptide that constitutes the core of cerebral plaques and the amyloid deposits in cerebral blood vessels in this disease. The Aβ protein is derived from a much larger transmembrane glycoprotein called amyloid precursor protein (APP) (Chapter 22).

Several other proteins have been found in amyloid deposits in a variety of clinical settings:

- **Transthyretin** (TTR) is a normal serum protein that binds and transports thyroxine and retinol, hence the name. Mutations in the gene encoding TTR may alter its structure, making the protein prone to misfolding and aggregation, and resistant to proteolysis. This leads to the formation of aggregates that deposit as amyloid. The resultant diseases are called familial amyloid polyneuropathies. TTR is also deposited in the heart of aged persons (senile systemic amyloidosis); in such cases the protein is structurally normal, but it accumulates at high concentrations. Some cases of familial amyloidosis are associated with deposits of mutant lysozyme.
- **β_2-Microglobulin,** a component of MHC class I molecules and a normal serum protein, has been identified as the amyloid fibril subunit (Aβ2m) in amyloidosis that complicates the course of patients on long-term hemodialysis. Aβ2m fibers are structurally similar to normal β2m protein. This protein is present in high concentrations in the serum of patients with renal disease and is retained in the circulation because it is not efficiently filtered through dialysis membranes. In some series, as many as 60% to 80% of patients on long-term dialysis developed amyloid deposits in the synovium, joints, and tendon sheaths.
- Amyloid deposits derived from diverse precursors such as hormones (procalcitonin) and keratin also have been reported.

Classification of Amyloidosis

Because a given biochemical form of amyloid (e.g., AA) may be associated with amyloid deposition in diverse clinical settings, a combined biochemical and clinical classification is followed for this discussion (Table 4–13). Amyloid may be systemic (generalized), involving several organ systems, or it may be localized, when deposits are limited to a single organ, such as the heart. On clinical grounds, the systemic, or generalized, pattern is subclassified into *primary amyloidosis* when associated with a monoclonal

Table 4–13 Classification of Amyloidosis

Clinicopathologic Category	Associated Disease(s)	Major Fibril Protein	Chemically Related Precursor Protein
Systemic (Generalized) Amyloidosis			
Immunocyte dyscrasias with amyloidosis (primary amyloidosis)	Multiple myeloma and other monoclonal plasma cell proliferations	AL	Immunoglobulin light chains, chiefly λ type
Reactive systemic amyloidosis (secondary amyloidosis)	Chronic inflammatory conditions	AA	SAA
Hemodialysis-associated amyloidosis	Chronic renal failure	$A\beta_2m$	β_2-Microglobulin
Hereditary Amyloidosis			
Familial Mediterranean fever		AA	SAA
Familial amyloidotic neuropathies (several types)		ATTR	Transthyretin
Systemic senile amyloidosis		ATTR	Transthyretin
Localized Amyloidosis			
Senile cerebral	Alzheimer disease	Aβ	APP
Endocrine			
Medullary carcinoma of thyroid		A Cal	Calcitonin
Islets of Langerhans	Type 2 diabetes	AIAPP	Islet amyloid peptide
Isolated atrial amyloidosis		AANF	Atrial natriuretic factor

plasma cell proliferation and *secondary amyloidosis* when it occurs as a complication of an underlying chronic inflammatory or tissue destructive process. Hereditary or familial amyloidosis constitutes a separate, albeit heterogeneous group, with several distinctive patterns of organ involvement.

Primary Amyloidosis: Immunocyte Dyscrasias with Amyloidosis

Amyloid in this category usually is systemic in distribution and is of the AL type. With approximately 3000 new cases each year in the United States, this is the most common form of amyloidosis. In some of these cases, there is a readily identifiable monoclonal plasma cell proliferation; best defined is the occurrence of systemic amyloidosis in 5% to 15% of patients with multiple myeloma, a plasma cell tumor characterized by multiple osteolytic lesions throughout the skeletal system (Chapter 11). The malignant plasma cells characteristically synthesize abnormal amounts of a single specific immunoglobulin (monoclonal gammopathy), producing an M (myeloma) protein spike on serum electrophoresis. In addition to the synthesis of whole immunoglobulin molecules, plasma cells also may synthesize and secrete either the λ or κ light chain, also known as Bence Jones proteins. By virtue of their small molecular size, these proteins frequently are also excreted in the urine. Almost all patients with myeloma who develop amyloidosis have Bence Jones proteins in the serum or urine, or both. However, amyloidosis develops in only 6% to 15% of patients with myeloma who have free light chains. Clearly, the presence of Bence Jones proteins, although necessary, is by itself not sufficient to produce amyloidosis. Other variables, such as the type of light chain produced and its catabolism, contribute to the "amyloidogenic potential" and influence the deposition of Bence Jones proteins.

The great majority of patients with AL amyloid do not have classic multiple myeloma or any other overt B cell neoplasm; such cases are nevertheless classified as primary amyloidosis because their clinical features derive from the effects of amyloid deposition without any other associated disease. In virtually all such cases, patients have a modest increase in the number of plasma cells in the bone marrow, and monoclonal immunoglobulins or free light chains can be found in the serum or urine. Clearly, these patients have an underlying monoclonal plasma cell proliferation in which production of an abnormal protein, rather than production of tumor masses, is the predominant manifestation.

Reactive Systemic Amyloidosis

The amyloid deposits in this pattern are systemic in distribution and are composed of AA protein. This category was previously referred to as secondary amyloidosis, because it is *secondary to an associated inflammatory condition.* In fact, the feature common to most cases of reactive systemic amyloidosis is chronic inflammation. Classically, tuberculosis, bronchiectasis, and chronic osteomyelitis were the most common causes; with the advent of effective antimicrobial therapies, reactive systemic amyloidosis is seen most frequently in the setting of chronic inflammation caused by autoimmune states (e.g., RA, ankylosing spondylitis, inflammatory bowel disease). Patients with RA are particularly prone to develop amyloidosis, with amyloid deposition seen in as many as 3% of RA cases. Chronic skin infections caused by "skin-popping" of narcotics are also associated with amyloid deposition. Finally, reactive systemic amyloidosis may also occur in association with tumors not derived from immune cells, the two most common being renal cell carcinoma and Hodgkin lymphoma.

Familial (Hereditary) Amyloidosis

A variety of familial forms of amyloidosis have been described; most are rare and occur in limited geographic areas. The best-characterized is an autosomal recessive condition called familial Mediterranean fever. This is a febrile disorder characterized by attacks of fever

accompanied by inflammation of serosal surfaces, including peritoneum, pleura, and synovial membrane. This disorder is encountered largely in persons of Armenian, Sephardic Jewish, and Arabic origins. It is associated with widespread tissue involvement indistinguishable from reactive systemic amyloidosis. The amyloid fibril proteins are made up of AA proteins, suggesting that this form of amyloidosis is related to the recurrent bouts of inflammation that characterize this disease. The gene for familial Mediterranean fever is called *pyrin* and encodes a protein that is a component of the inflammasome (Chapter 2). Patients have gain-of-function mutations in *pyrin* that result in constitutive overproduction of the pro-inflammatory cytokine IL-1 and persistent inflammation.

In contrast with familial Mediterranean fever, a group of autosomal dominant familial disorders is characterized by deposition of amyloid predominantly in the peripheral and autonomic nerves. These familial amyloidotic polyneuropathies have been described in kindreds in different parts of the world—for example, in Portugal, Japan, Sweden, and the United States. As mentioned previously, the fibrils in these familial polyneuropathies are made up of mutant forms of transthyretin (ATTRs).

Localized Amyloidosis

Sometimes deposition of amyloid is limited to a single organ or tissue without involvement of any other site in the body. The deposits may produce grossly detectable nodular masses or be evident only on microscopic examination. Nodular (tumor-forming) deposits of amyloid are most often encountered in the lung, larynx, skin, urinary bladder, tongue, and the region about the eye. Frequently, there are infiltrates of lymphocytes and plasma cells in the periphery of these amyloid masses, raising the question of whether the mononuclear infiltrate is a response to the deposition of amyloid or instead is responsible for it. At least in some cases, the amyloid consists of AL protein and may therefore represent a localized form of plasma cell–derived amyloid.

Endocrine Amyloid

Microscopic deposits of localized amyloid may be found in certain endocrine tumors, such as medullary carcinoma of the thyroid gland, islet tumors of the pancreas, pheochromocytomas, and undifferentiated carcinomas of the stomach, as well as in the islets of Langerhans in patients with type 2 diabetes mellitus. In these settings, the amyloidogenic proteins seem to be derived either from polypeptide hormones (medullary carcinoma) or from unique proteins (e.g., islet amyloid polypeptide).

Amyloid of Aging

Several well-documented forms of amyloid deposition occur with aging. Senile systemic amyloidosis refers to the systemic deposition of amyloid in elderly persons (usually in their 70s and 80s). Because of the dominant involvement and related dysfunction of the heart (typically manifesting as a restrictive cardiomyopathy and arrhythmias), this form also is called *senile cardiac amyloidosis*. The amyloid in this form is composed of normal transthyretin. In addition, another form typically affecting only the heart results from the deposition of a mutant form of TTR. Approximately 4% of the black population in the United States are carriers of the mutant allele, and cardiomyopathy has been identified in both homozygous and heterozygous patients.

MORPHOLOGY

There are no consistent or distinctive patterns of organ or tissue distribution of amyloid deposits in any of the categories cited. Nonetheless, a few generalizations can be made. In amyloidosis secondary to chronic inflammatory disorders, kidneys, liver, spleen, lymph nodes, adrenals, and thyroid, as well as many other tissues, typically are affected. Although primary (AL) amyloidosis cannot reliably be distinguished from the secondary form by its organ distribution, it more often involves the heart, gastrointestinal tract, respiratory tract, peripheral nerves, skin, and tongue. However, the same organs affected by reactive systemic amyloidosis (secondary amyloidosis), including kidneys, liver, and spleen, also may contain deposits in the immunocyte-associated form of the disease. The localization of amyloid deposits in the hereditary syndromes is varied. In familial Mediterranean fever, the amyloidosis may be widespread, involving the kidneys, blood vessels, spleen, respiratory tract, and (rarely) liver. The localization of amyloid in the remaining hereditary syndromes can be inferred from the designation of these entities.

Whatever the clinical disorder, the amyloidosis may or may not be apparent grossly. Often small amounts are not recognized until the surface of the cut organ is painted with iodine and sulfuric acid. This yields mahogany brown staining of the amyloid deposits. When amyloid accumulates in larger amounts, the organ frequently is enlarged and the tissue typically appears gray with a waxy, firm consistency. **On histologic examination, the amyloid deposition is always extracellular and begins between cells,** often closely adjacent to basement membranes. As the amyloid accumulates, it encroaches on the cells, in time surrounding and destroying them. In the AL form, perivascular and vascular localizations are common.

The histologic diagnosis of amyloid is based almost entirely on its staining characteristics. The most commonly used staining technique uses the dye Congo red, which under ordinary light imparts a pink or red color to amyloid deposits. Under polarized light the Congo red–stained amyloid shows so-called apple-green birefringence (Fig. 4–33). This reaction is shared by all forms of amyloid and is caused by the crossed β-pleated configuration of amyloid fibrils. Confirmation can be obtained by electron microscopy, which reveals amorphous nonoriented thin fibrils. AA, AL, and ATTR types of amyloid also can be distinguished from one another by specific immunohistochemical staining.

Because the pattern of organ involvement in different clinical forms of amyloidosis is variable, each of the major organ involvements is described separately.

Kidney. Amyloidosis of the kidney is the most common and most serious feature of the disease. Grossly, the kidney may appear unchanged, or it may be abnormally large, pale, gray, and firm; in long-standing cases, the kidney may be reduced in size. Microscopically, the **amyloid deposits are found principally in the glomeruli,** but they also are present in the interstitial peritubular tissue as well as in the walls of the blood vessels. The glomerulus first develops focal deposits

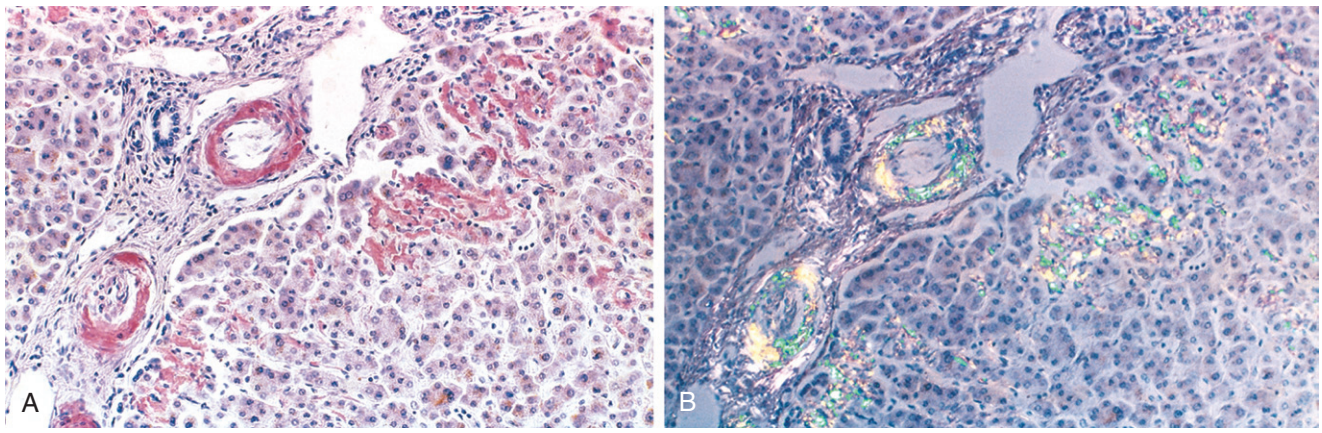


Figure 4–33 Amyloidosis: hepatic involvement. **A,** Staining of a section of the liver with Congo red reveals pink-red deposits of amyloid in the walls of blood vessels and along sinusoids. **B,** Note the yellow-green birefringence of the deposits when observed under the polarizing microscope.
(Courtesy of Dr. Trace Worrell and Sandy Hinton, Department of Pathology, University of Texas Southwestern Medical School, Dallas, Texas.)

within the mesangial matrix and diffuse or nodular thickenings of the basement membranes of the capillary loops. With progression, the deposition encroaches on the capillary lumina and eventually leads to total obliteration of the vascular tuft (Fig. 4–34, *A*). The interstitial peritubular deposits frequently are associated with the appearance of amorphous pink casts within the tubular lumens, presumably of a proteinaceous nature. Amyloid deposits may develop in the walls of blood vessels of all sizes, often causing marked vascular narrowing.

Spleen. Amyloidosis of the spleen often causes moderate or even marked enlargement (200 to 800 gm). For obscure reasons, either of two patterns may develop. The deposits may be virtually limited to the splenic follicles, producing tapioca-like granules on gross examination ("sago spleen"), or the amyloidosis may principally involve the splenic sinuses, eventually extending to the splenic pulp, with formation of large, sheetlike deposits ("lardaceous spleen"). In both patterns, the spleen is firm in consistency. The presence of blood in splenic sinuses usually imparts a reddish color to the waxy, friable deposits.

Liver. Amyloidosis of the liver may cause massive enlargement (as much as 9000 gm). In such advanced cases, the liver is extremely pale, grayish, and waxy on both the external surface and the cut section. Histologic analysis shows that **amyloid deposits first appear in the space of Disse** and then progressively enlarge to encroach on the adjacent hepatic parenchyma and sinusoids (Fig. 4–33). The trapped liver cells undergo compression atrophy and are eventually replaced by sheets of amyloid; remarkably, normal liver function may be preserved even in the setting of severe involvement.

Heart. Amyloidosis of the heart may occur either as isolated organ involvement or as part of a systemic distribution. When accompanied by systemic involvement, it is usually of the AL form. The isolated form (senile amyloidosis) usually is confined to older persons. The deposits may not be evident on gross examination, or they may cause minimal to moderate cardiac enlargement. The most characteristic gross findings are gray-pink, dewdrop-like subendocardial elevations, particularly evident in the atrial chambers. On histologic examination, deposits typically are found throughout the myocardium, beginning **between myocardial fibers** and eventually causing their pressure atrophy (Fig. 4–34, *B*).

Other Organs. Amyloidosis of other organs generally is encountered in systemic disease. The adrenals, thyroid, and

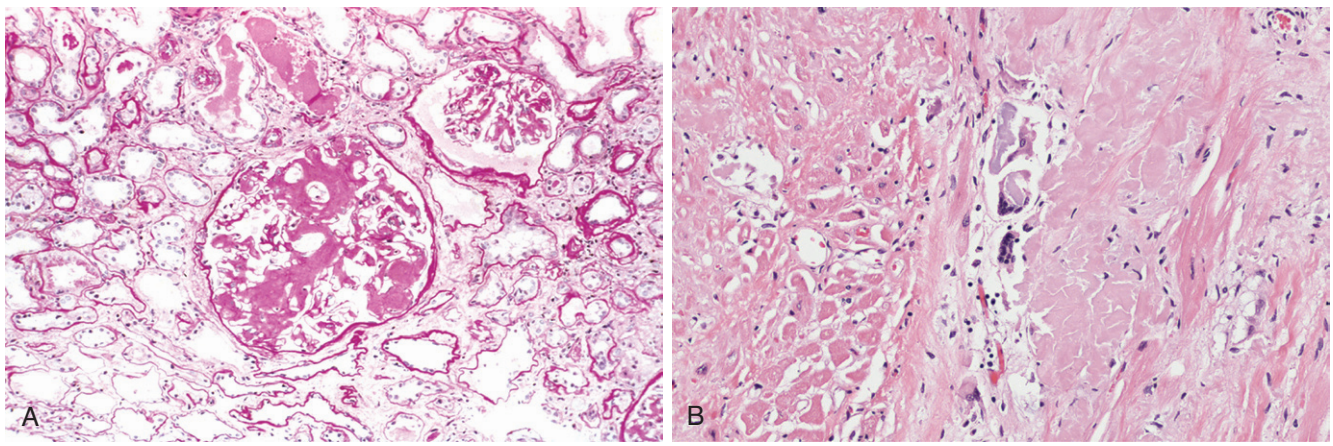


Figure 4–34 Amyloidosis: renal and cardiac involvement. **A,** Amyloidosis of the kidney. The glomerular architecture is almost totally obliterated by the massive accumulation of amyloid. **B,** Cardiac amyloidosis. The atrophic myocardial fibers are separated by structureless, pink-staining amyloid.

pituitary are common sites of involvement. In such cases as well, the amyloid deposition begins in relation to stromal and endothelial cells and progressively encroaches on the parenchymal cells. Surprisingly large amounts of amyloid may be present in any of these endocrine glands without apparent disturbance of function. In the gastrointestinal tract, a relatively favored site for deposition, amyloid may be found at all levels, sometimes producing tumorous masses that must be distinguished from neoplasms. Nodular depositions in the tongue may produce **macroglossia.** On the basis of the frequent involvement of the gastrointestinal tract in systemic cases, gingival, intestinal, and rectal biopsies serve in the diagnosis of suspected cases. Deposition of β_2-microglobulin amyloid in patients receiving long-term dialysis occurs most commonly in the **carpal ligaments of the wrist,** resulting in compression of the median nerve (leading to carpal tunnel syndrome).

Clinical Course

Amyloidosis may be an unsuspected finding at autopsy in a patient who has no apparent related clinical manifestations, or it may be responsible for serious clinical dysfunction and even death. The clinical course depends on the particular sites or organs affected and the severity of the involvement. Nonspecific complaints such as weakness, fatigue, and weight loss are the most common presenting manifestations. Later in the course, amyloidosis tends to manifest in one of several ways: by renal disease, hepatomegaly, splenomegaly, or cardiac abnormalities. Renal involvement giving rise to severe proteinuria (nephrotic syndrome) (Chapter 13) often is the major cause of symptoms in reactive systemic amyloidosis. Progression of the renal disease may lead to renal failure, which is an important cause of death in amyloidosis. The hepatosplenomegaly rarely causes significant clinical dysfunction, but it may be the presenting finding. Cardiac amyloidosis may manifest as conduction disturbances or as restrictive cardiomyopathy (Chapter 10). Cardiac arrhythmias are an important cause of death in cardiac amyloidosis. In one large series, 40% of the patients with AL amyloid died of cardiac disease.

The diagnosis of amyloidosis may be suspected from the clinical signs and symptoms and from some of the findings mentioned; however, more specific tests must often be done for definitive diagnosis. Biopsy and subsequent Congo red staining is the most important tool in the diagnosis of amyloidosis. In general, biopsy is taken from the organ suspected to be involved. For example, renal biopsy is useful in the presence of urinary abnormalities. Rectal and gingival biopsy specimens contain amyloid in as many as 75% of cases with generalized amyloidosis. Examination of abdominal fat aspirates stained with Congo red is a simple, low-risk method. In suspected cases of AL amyloidosis, serum and urinary protein electrophoresis and immunoelectrophoresis should be performed. Bone marrow examination in such cases usually shows plasmacytosis, even if skeletal lesions of multiple myeloma are not present. Proteomic analysis of affected tissue is now being widely used for detection of small amounts of amyloid (from fat aspirates) and for definitive identification of the type of amyloid.

The outlook for patients with generalized amyloidosis is poor, with the mean survival time after diagnosis ranging from 1 to 3 years. In AA amyloidosis, the prognosis depends to some extent on the control of the underlying condition. Patients with myeloma-associated amyloidosis have a poorer prognosis, although they may respond to cytotoxic drugs used to treat the underlying disorder. Resorption of amyloid after treatment of the associated condition has been reported, but this is a rare occurrence.

SUMMARY

Amyloidosis

- Amyloidosis is a disorder characterized by the extracellular deposits of misfolded proteins that aggregate to form insoluble fibrils.
- The deposition of these proteins may result from excessive production of proteins that are prone to misfolding and aggregation; mutations that produce proteins that cannot fold properly and tend to aggregate; or defective or incomplete proteolytic degradation of extracellular proteins.
- Amyloidosis may be localized or systemic. It is seen in association with a variety of primary disorders, including monoclonal plasma cell proliferations (in which the amyloid deposits consist of immunoglobulin light chains); chronic inflammatory diseases such as RA (deposits of amyloid A protein, derived from an acute-phase protein produced in inflammation); Alzheimer disease (amyloid B protein); familial conditions in which the amyloid deposits consist of mutants of normal proteins (e.g., transthyretin in familial amyloid polyneuropathies); amyloidosis associated with dialysis (deposits of β_2-microglobulin, whose clearance is defective).
- Amyloid deposits cause tissue injury and impair normal function by causing pressure on cells and tissues. They do not evoke an inflammatory response.

BIBLIOGRAPHY

Banchereau J, Pascual V: Type I interferon in systemic lupus erythematosus and other autoimmune diseases. Immunity 25:383, 2006. *[A review of the recently discovered role of interferons in SLE and other autoimmune diseases, and the potential for targeting this family of cytokines for therapy.]*

Campbell DJ, Koch MA: Phenotypic and functional specialization of FoxP3+ regulatory T cells. Nat Rev Immunol 11:119, 2011. *[A current review of the properties and functions of regulatory T cells.]*

Chervonsky A: Influence of microbial environment on autoimmunity. Nat Immunol 11:28, 2010. *[A summary of the role of microbes and other environmental factors in the development of autoimmunity.]*

Cunningham-Rundles C, Ponda PP: Molecular defects in T- and B-cell primary immunodeficiency diseases. Nat Rev Immunol 5:880, 2006. *[Excellent, up-to-date review of primary immunodeficiencies.]*

Davidson A, Diamond B: Autoimmune diseases. N Engl J Med 345:340, 2001. *[A readable overview of the etiology, pathogenesis, and therapy for autoimmune diseases.]*

Douek DC, Roederer M, Koup RA: Emerging concepts in the immunopathogenesis of AIDS. Annu Rev Med 60:471, 2009. *[A balanced discussion of the pathogenesis of AIDS, and the still unresolved issues.]*

Fairhurst AM, Wandstrat AE, Wakeland EK: Systemic lupus erythematosus: multiple immunological phenotypes in a complex genetic disease. Adv Immunol 92:1, 2006. *[A comprehensive review of the*

genetics of SLE in humans and animal models, and the pathogenic mechanisms deduced from identification of susceptibility genes.]

Fischer A: Human primary immunodeficiency diseases. Immunity 28:835, 2008. *[An excellent summary of primary immunodeficiencies affecting the innate and adaptive immune systems.]*

Galli SJ: The development of allergic inflammation. Nature 454:445, 2008. *[An excellent review of the mechanisms of inflammation in allergic diseases.]*

Gonzalez-Scarano F, Martin-Garcia J: The neuropathogenesis of AIDS. Nat Rev Immunol 5:69, 2005. *[A discussion of the pathogenesis of HIV-associated dementia.]*

Goodnow CC: Multistep pathogenesis of autoimmune disease. Cell 130:25, 2007. *[An excellent discussion of the checkpoints that prevent autoimmunity and why these might fail.]*

Jancar S, Sanchez Crespo M: Immune complex–mediated tissue injury: a multistep paradigm. Trends Immunol 26:48, 2005. *[A summary of the mechanisms of immune complex–mediated tissue injury.]*

Katsumoto TR, Whitfield ML, Connolly MK: The pathogenesis of systemic sclerosis. Annu Rev Pathol 6:509, 2011. *[An excellent review of the pathogenesis of systemic sclerosis, and the many unanswered questions.]*

Kay AB: Allergy and allergic diseases. First of two parts. N Engl J Med 344:30, 2001. *[Excellent two-part review of the mechanisms and manifestations of immediate hypersensitivity.]*

Kyewski B, Klein L: A central role for central tolerance. Annu Rev Immunol 24:571, 2006. *[A discussion of the mechanisms of central tolerance, with a focus on T cells.]*

Merlini G, Bellotti V: Molecular mechanisms of amyloidosis. N Engl J Med 349:583, 2003. *[Good review of the general pathogenic mechanisms in various forms of systemic amyloidosis.]*

Mitchell RN: Graft vascular disease: immune response meets the vessel wall. Annu Rev Pathol 4:19, 2009. *[A review of the mechanisms that lead to vascular disease in chronic graft rejection.]*

Moir S, Chun TW, Fauci AS: Pathogenic mechanisms of HIV disease. Annu Rev Pathol 6:223, 2011. *[A discussion of current concepts of the mechanisms by which HIV causes immunodeficiency.]*

Mueller DL: Mechanisms maintaining peripheral tolerance. Nat Immunol 11:21, 2010. *[A discussion of the mechanisms of peripheral tolerance, with an emphasis on T cells.]*

Nankivell BJ, Alexander SI: Rejection of the kidney allograft. N Engl J Med 363:1451, 2010. *[Good review of the mechanisms of recognition and rejection of allografts and the development of new strategies for treating rejection.]*

Notarangelo LD: Primary immunodeficiencies. J Allergy Clin Immunol 125:S182, 2010. *[An update on congenital immunodeficiencies and their molecular basis.]*

O'Shea JJ, Paul WE: Mechanisms underlying lineage commitment and plasticity of helper CD4+ T cells. Science 327:1098, 2010. *[An excellent review of the development and functions of helper T cell subsets, and the uncertainties in the field.]*

Palmer MT, Weaver CT: Autoimmunity: increasing suspects in the CD4+ T cell lineage. Nat Immunol 11:36, 2010. *[A thoughtful discussion of the central role of CD4+ T cells in the pathogenesis of autoimmune diseases.]*

Pepys MB: Amyloidosis. Annu Rev Med 57:223, 2006. *[An excellent review of the pathogenesis, clinical features and therapeutic approaches in amyloidosis.]*

Rahman A, Isenberg DA: Systemic lupus erythematosus. New Engl J Med 358:929, 2008. *[An excellent review of the pathogenesis and genetics of SLE.]*

Sakaguchi S, Miyara M, Costantino CM, Hafler DA: FOXP3+ regulatory T cells in the human immune system. Nat Rev Immunol 10:490, 2010. *[An excellent discussion of the properties and role of regulatory T cells in humans.]*

Sakaguchi S, Yamaguchi T, Nomura T, Ono M: Regulatory T cells and immune tolerance. Cell 133:775, 2008. *[An excellent review of the properties and functions of regulatory T cells.]*

Stone J, Zen Y, Deshpande V: IgG4-related disease. New Engl J Med, in press, 2011. *[A comprehensive discussion of the features and likely pathogenesis of this recently recognized entity.]*

Vercelli D: Discovering susceptibility genes for asthma and allergy. Nat Rev Immunol 8:169, 2008. *[A summary of the known susceptibility genes for asthma and other allergic diseases.]*

Voulgarelis M, Tzioufas AG: Pathogenetic mechanisms in the initiation and perpetuation of Sjögren's syndrome. Nat Rev Rheumatol 6:529, 2010. *[A good discussion of what is known and not known about the pathogenesis of Sjögren's syndrome.]*

Zenewicz L, Abraham C, Flavell RA, Cho J: Unraveling the genetics of autoimmunity. Cell 140:791, 2010. *[An update on susceptibility genes for autoimmune diseases, how these are identified, and their significance.]*

This page intentionally left blank

See Targeted Therapy available online at studentconsult.com

CHAPTER 5

Neoplasia

CHAPTER CONTENTS

Cancer is the second leading cause of death in the United States; only cardiovascular diseases exact a higher toll. Even more agonizing than the associated mortality is the emotional and physical suffering inflicted by neoplasms. Patients and the public often ask, "When will there be a cure for cancer?" The answer to this simple question is difficult, because cancer is not one disease but many disorders that share a profound growth dysregulation. Some cancers, such as Hodgkin lymphomas, are highly curable, whereas others, such as cancer of the pancreas, are virtually always fatal. The only hope for controlling cancer lies in learning more about its pathogenesis, and great strides have been made in understanding the molecular basis of cancer. This chapter deals with the basic biology of neoplasia—the nature of benign and malignant neoplasms and the molecular basis of neoplastic transformation. The host response to tumors and the clinical features of neoplasia also are discussed. Before we discuss the features of cancer cells and the mechanisms of carcinogenesis, it is useful to summarize the fundamental and shared characteristics of cancers:

- *Cancer is a genetic disorder* caused by DNA mutations that are (for the most part) acquired spontaneously or induced by environmental insults. In addition, cancers frequently show epigenetic changes, such as focal increases in DNA methylation and alterations in histone modifications, which may themselves stem from acquired mutations in genes that regulate such modifications. These genetic and epigenetic changes alter the expression or function of key genes that regulate fundamental cellular processes, such as growth, survival, and senescence.
- *These genetic alterations are heritable, being passed to daughter cells upon cell division*. As a result, *cells harboring these alterations are subject to darwinian selection* (survival of the fittest, arguably the most important scientific concept yet conceived), with cells bearing mutations that provide them with growth or survival advantages outcompeting their neighbors and thus coming to dominate the population. Darwinian selection also plays a role in the progression and recurrence of cancers, as discussed in more detail later. Because the selective advantages are conferred on a single cell that ultimately gives rise to the tumor, all tumors are *clonal* (i.e., the progeny of one cell).
- *Accumulation of mutations gives rise to a set of properties that have been called hallmarks of cancer*. These include (1) self-sufficiency in growth signals whereby the growth of cancers becomes autonomous and is unregulated by physiologic cues; (2) lack of response to growth inhibitory signals that control non-neoplastic cellular

proliferations such as hyperplasias; (3) evasion of cell death, allowing cancer cells to survive under conditions that induce apoptosis in normal cells; (4) limitless replicative potential, thus making cancer cells immortal; (5) development of angiogenesis to sustain the growth of cancer cells; (6) ability to invade local tissues and spread to distant sites; (7) reprogramming of metabolic pathways—specifically, a switch to aerobic glycolysis even when there is abundant oxygen; and (8) ability to evade the immune system. The genetic alterations that give rise to these hallmarks of cancers are sustained and enabled by the development of genomic instability, adding fuel to the fire. The molecular underpinnings of these hallmarks are discussed in detail in a later section.

Understanding the cellular and molecular abnormalities in cancer cells is leading to a revolution in the treatment of cancer founded on basic research, and is one of the emerging triumphs of biomedical science.

NOMENCLATURE

Neoplasia literally means "new growth." Neoplastic cells are said to be *transformed* because they continue to replicate, apparently oblivious to the regulatory influences that control normal cell growth. Neoplasms therefore enjoy a certain degree of autonomy and tend to increase in size regardless of their local environment. Their autonomy is by no means complete, however. Some neoplasms require endocrine support, and such dependencies sometimes can be exploited therapeutically. All neoplasms depend on the host for their nutrition and blood supply.

In common medical usage, a neoplasm often is referred to as a *tumor,* and the study of tumors is called *oncology* (from *oncos,* "tumor," and *logos,* "study of"). Among tumors, the division of neoplasms into benign and malignant categories is based on a judgment of a tumor's potential clinical behavior.

- A tumor is said to be *benign* when its microscopic and gross characteristics are considered to be relatively innocent, implying that it will remain localized and is amenable to local surgical removal; the patient generally survives. Of note, however, benign tumors can produce more than localized lumps, and sometimes they are responsible for serious disease.
- Malignant tumors are collectively referred to as *cancers,* derived from the Latin word for "crab"—that is, they adhere to any part that they seize in an obstinate manner, similar to a crab's behavior. *Malignant,* as applied to a neoplasm, implies that the lesion can invade and destroy adjacent structures and spread to distant sites (metastasize) to cause death. Not all cancers pursue so deadly a course. The most aggressive are also some of the most curable, but the designation *malignant* constitutes a red flag.

All tumors, benign and malignant, have two basic components: (1) the *parenchyma,* made up of transformed or neoplastic cells, and (2) the supporting, host-derived, nonneoplastic *stroma,* made up of connective tissue, blood vessels, and host-derived inflammatory cells. The parenchyma of the neoplasm largely determines its biologic behavior, and it is this component from which the tumor derives its name. The stroma is crucial to the growth of the neoplasm, since it carries the blood supply and provides support for the growth of parenchymal cells. Although the biologic behavior of tumors largely reflects the behavior of the parenchymal cells, there has been a growing realization that stromal cells and neoplastic cells carry on a two-way conversation that influences the growth of the tumor.

Benign Tumors

In general, benign tumors are designated by attaching the suffix *-oma* to the cell type from which the tumor arises. A benign tumor arising in fibrous tissue is a *fibroma*; a benign cartilaginous tumor is a *chondroma.* The nomenclature of benign epithelial tumors is more complex. They are classified sometimes on the basis of their microscopic pattern and sometimes on the basis of their macroscopic pattern. Others are classified by their cells of origin.

For instance, the term *adenoma* is generally applied to benign epithelial neoplasms producing gland patterns and to neoplasms derived from glands but not necessarily exhibiting glandular patterns. A benign epithelial neoplasm arising from renal tubule cells and growing in glandlike patterns is termed an adenoma, as is a mass of benign epithelial cells that produces no glandular patterns but has its origin in the adrenal cortex. *Papillomas* are benign epithelial neoplasms, growing on any surface, that produce microscopic or macroscopic finger-like fronds. A *polyp* is a mass that projects above a mucosal surface, as in the gut, to form a macroscopically visible structure (Fig. 5–1). Although this term commonly is used for benign tumors, some malignant tumors also may grow as polyps, whereas other polyps (such as nasal polyps) are not neoplastic but inflammatory in origin. *Cystadenomas* are hollow cystic masses that typically arise in the ovary.

Malignant Tumors

The nomenclature of malignant tumors essentially follows that of benign tumors, with certain additions and exceptions.

- Malignant neoplasms arising in "solid" mesenchymal tissues or its derivatives are called *sarcomas*, whereas those arising from the mesenchymal cells of the blood are called leukemias or lymphomas. Sarcomas are designated by the cell type of which they are composed, which is presumably their cell of origin. Thus, a cancer of fibrous tissue origin is a *fibro*sarcoma, and a malignant neoplasm composed of chondrocytes is a *chondro*sarcoma.
- While the epithelia of the body are derived from all three germ cell layers, malignant neoplasms of epithelial cells are called *carcinomas* regardless of the tissue of origin. Thus, a malignant neoplasm arising in the renal tubular epithelium (mesoderm) is a carcinoma, as are the cancers arising in the skin (ectoderm) and lining epithelium of the gut (endoderm). Furthermore, mesoderm may give rise to carcinomas (epithelial), sarcomas (mesenchymal), and hematolymphoid tumors (leukemias and lymphomas).
- Carcinomas are subdivided further. Carcinomas that grow in a glandular pattern are called *adenocarcinomas,*

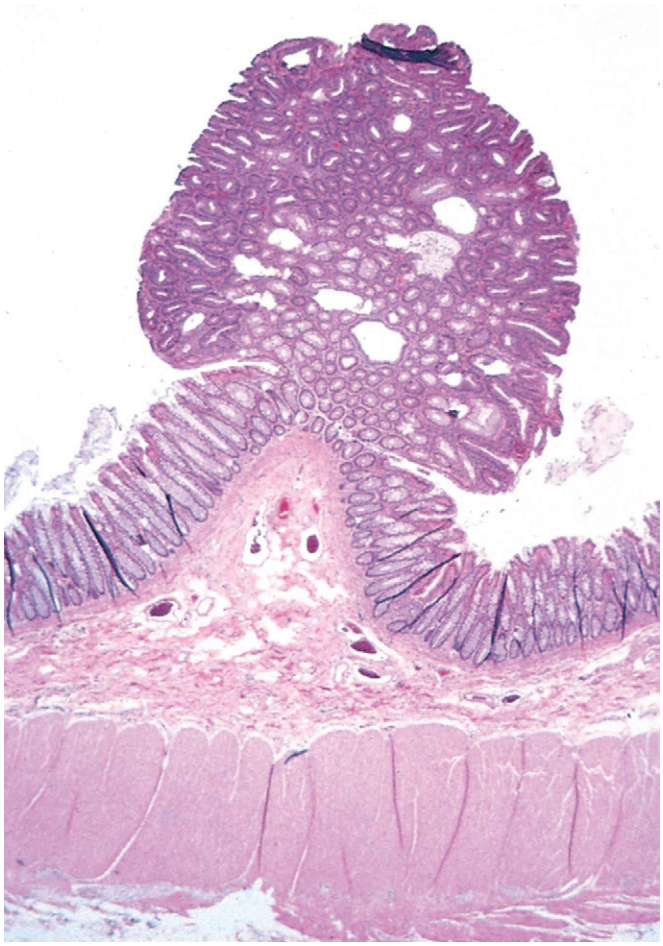


Figure 5–1 Colonic polyp. This glandular tumor (adenoma) is seen projecting into the colonic lumen. The polyp is attached to the mucosa by a distinct stalk.

and those that produce squamous cells are called *squamous cell carcinomas.* Sometimes the tissue or organ of origin can be identified, as in the designation of renal cell adenocarcinoma. Sometimes the tumor shows little or no differentiation and must be called *poorly differentiated or undifferentiated carcinoma.*

The transformed cells in a neoplasm, whether benign or malignant, often resemble each other, as though all had been derived from a single progenitor, consistent with the monoclonal origin of tumors. In some unusual instances, however, the tumor cells undergo *divergent differentiation,* creating so-called *mixed tumors.* The best example is mixed tumor of salivary gland. These tumors have obvious epithelial components dispersed throughout a fibromyxoid stroma, sometimes harboring islands of cartilage or bone (Fig. 5–2). All of these diverse elements are thought to derive from epithelial cells or myoepithelial cells, or both, and the preferred designation for these neoplasms is *pleomorphic adenoma.* Fibroadenoma of the female breast is another common mixed tumor. This benign tumor contains a mixture of proliferating ductal elements (adenoma) embedded in a loose fibrous tissue (fibroma). Although only the fibrous component is neoplastic, the term *fibroadenoma* remains in common usage.

Teratoma is a special type of mixed tumor that contains recognizable mature or immature cells or tissues representative of more than one germ cell layer and sometimes all three. Teratomas originate from totipotential germ cells such as those normally present in the ovary and testis and sometimes abnormally present in sequestered midline embryonic rests. Germ cells have the capacity to differentiate into any of the cell types found in the adult body; not surprisingly, therefore, they may give rise to neoplasms that mimic, in helter-skelter fashion, bits of bone, epithelium, muscle, fat, nerve, and other tissues.

The specific names of the more common forms of neoplasms are presented in Table 5–1. Some glaring inconsistencies may be noted. For example, the terms *lymphoma, mesothelioma, melanoma,* and *seminoma* are used for malignant neoplasms. Unfortunately for students, these exceptions are firmly entrenched in medical terminology.

There are other instances of confusing terminology:

- *Hamartoma* is a mass of disorganized tissue indigenous to the particular site. Histopathologic examination may show a mass of mature but disorganized hepatic cells, blood vessels, and possibly bile ducts within the liver, or a nodule in the lung containing islands of cartilage, bronchi, and blood vessels. Hamartomas have traditionally been considered developmental malformations, but some genetic studies have shown the presence of acquired translocations, suggesting a neoplastic origin.
- *Choristoma* is a congenital anomaly consisting of a heterotopic rest of cells. For example, a small nodule of well-developed and normally organized pancreatic tissue may be found in the submucosa of the stomach, duodenum, or small intestine. This heterotopic rest may be replete with islets of Langerhans and exocrine glands. The designation *-oma,* connoting a neoplasm, imparts to the heterotopic rest a gravity far beyond its usual trivial significance.

Although the terminology of neoplasms is regrettably not simple, a firm grasp of the nomenclature is important because it is the language by which the nature and significance of tumors are categorized.

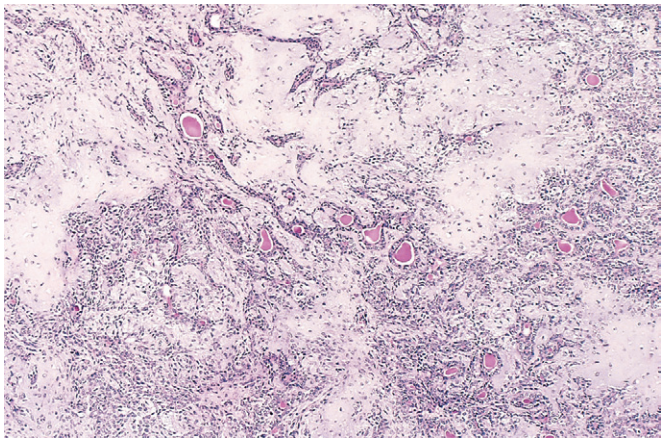


Figure 5–2 Mixed tumor of the parotid gland contains epithelial cells forming ducts and myxoid stroma that resembles cartilage.
(Courtesy of Dr. Trace Worrell, Department of Pathology, University of Texas Southwestern Medical School, Dallas, Texas.)

Table 5–1 Nomenclature of Tumors

Tissue of Origin	Benign	Malignant
Composed of One Parenchymal Cell Type		
Connective tissue and derivatives	Fibroma	Fibrosarcoma
	Lipoma	Liposarcoma
	Chondroma	Chondrosarcoma
	Osteoma	Osteogenic sarcoma
Endothelial and related tissues		
Blood vessels	Hemangioma	Angiosarcoma
Lymph vessels	Lymphangioma	Lymphangiosarcoma
Mesothelium		Mesothelioma
Brain coverings	Meningioma	Invasive meningioma
Blood cells and related cells		
Hematopoietic cells		Leukemias
Lymphoid tissue		Lymphomas
Muscle		
Smooth	Leiomyoma	Leiomyosarcoma
Striated	Rhabdomyoma	Rhabdomyosarcoma
Tumors of epithelial origin		
Stratified squamous	Squamous cell papilloma	Squamous cell or epidermoid carcinoma
Basal cells of skin or adnexa		Basal cell carcinoma
Epithelial lining of glands or ducts	Adenoma	Adenocarcinoma
	Papilloma	Papillary carcinomas
	Cystadenoma	Cystadenocarcinoma
Respiratory passages	Bronchial adenoma	Bronchogenic carcinoma
Renal epithelium	Renal tubular adenoma	Renal cell carcinoma
Liver cells	Liver cell adenoma	Hepatocellular carcinoma
Urinary tract epithelium (transitional)	Urothelial papilloma	Urothelial carcinoma
Placental epithelium	Hydatidiform mole	Choriocarcinoma
Testicular epithelium (germ cells)		Seminoma
		Embryonal carcinoma
Tumors of melanocytes	Nevus	Malignant melanoma
More Than One Neoplastic Cell Type—Mixed Tumors, Usually Derived from One Germ Cell Layer		
Salivary glands	Pleomorphic adenoma (mixed tumor of salivary gland)	Malignant mixed tumor of salivary gland
Renal anlage		Wilms tumor
More Than One Neoplastic Cell Type Derived from More Than One Germ Cell Layer—Teratogenous		
Totipotential cells in gonads or in embryonic rests	Mature teratoma, dermoid cyst	Immature teratoma, teratocarcinoma

CHARACTERISTICS OF BENIGN AND MALIGNANT NEOPLASMS

Nothing is more important to the patient with a tumor than being told: "It is benign." In general, benign tumors appear to be genetically "simple," harboring fewer mutations than cancers, and genetically stable, changing little in genotype over time. The latter feature probably explains why benign tumors such as lipomas and leiomyomas transform to malignancies rarely, if at all. In practice, the determination of benign versus malignant is made with remarkable accuracy using long-established clinical and anatomic criteria, but some neoplasms defy easy characterization. Certain features may indicate innocence, and others may indicate malignancy. Such problems are not the rule, however, and *there are four fundamental features by which benign and malignant tumors can be distinguished: differentiation and anaplasia, rate of growth, local invasion, and metastasis.*

Differentiation and Anaplasia

Differentiation and anaplasia are characteristics seen only in the parenchymal cells that constitute the transformed elements of neoplasms. The differentiation of parenchymal tumor cells refers to the extent to which they resemble their normal forebears morphologically and functionally.

- Benign neoplasms are composed of well-differentiated cells that closely resemble their normal counterparts. A lipoma is made up of mature fat cells laden with cytoplasmic lipid vacuoles, and a chondroma is made up of mature cartilage cells that synthesize their usual cartilaginous matrix—evidence of morphologic and functional differentiation. In well-differentiated benign tumors, mitoses are usually rare and are of normal configuration.
- Malignant neoplasms are characterized by a wide range of parenchymal cell differentiation, from surprisingly well differentiated (Fig. 5–3) to completely

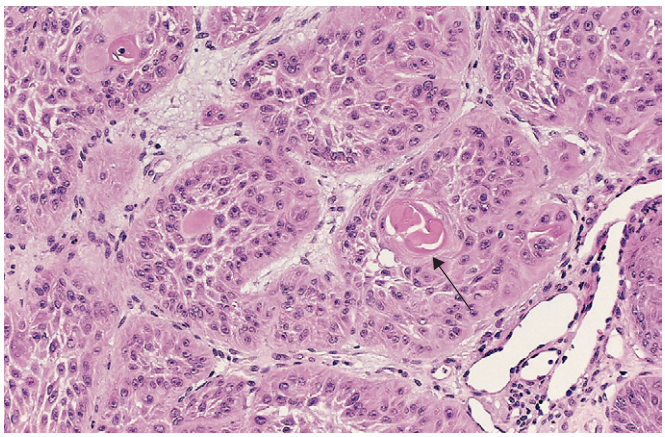


Figure 5–3 Well-differentiated squamous cell carcinoma of the skin. The tumor cells are strikingly similar to normal squamous epithelial cells, with intercellular bridges and nests of keratin *(arrow)*.
(Courtesy of Dr. Trace Worrell, Department of Pathology, University of Texas Southwestern Medical School, Dallas, Texas.)

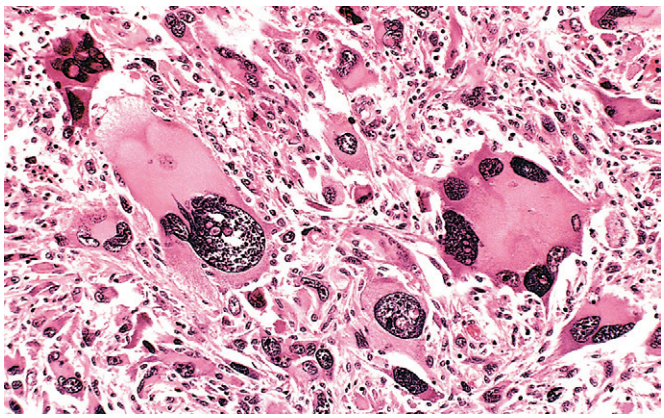


Figure 5–4 Anaplastic tumor of the skeletal muscle (rhabdomyosarcoma). Note the marked cellular and nuclear pleomorphism, hyperchromatic nuclei, and tumor giant cells.
(Courtesy of Dr. Trace Worrell, Department of Pathology, University of Texas Southwestern Medical School, Dallas, Texas.)

undifferentiated. For example, well-differentiated adenocarcinomas of the thyroid may contain normal-appearing follicles. Such tumors sometimes may be difficult to distinguish from benign proliferations. Between the two extremes lie tumors loosely referred to as *moderately well differentiated.* The stroma carrying the blood supply is crucial to the growth of tumors but does not aid in the separation of benign from malignant ones. The amount of stromal connective tissue does determine, however, the consistency of a neoplasm. Certain cancers induce a dense, abundant fibrous stroma (desmoplasia), making them hard, so-called scirrhous tumors.

- Malignant neoplasms that are composed of undifferentiated cells are said to be *anaplastic.* Lack of differentiation, or anaplasia, is considered a hallmark of malignancy. The term *anaplasia* literally means "backward formation"—implying dedifferentiation, or loss of the structural and functional differentiation of normal cells. It is now known, however, that at least some cancers arise from stem cells in tissues; in these tumors, failure of differentiation, rather than dedifferentiation of specialized cells, accounts for their undifferentiated appearance. Recent studies also indicate that in some cases, dedifferentiation of apparently mature cells does occur during carcinogenesis. Anaplastic cells display marked *pleomorphism* (i.e., variation in size and shape) (Fig. 5–4). Often the *nuclei are extremely hyperchromatic* (dark-staining) and large resulting in an increased nuclear-to-cytoplasmic ratio that may approach 1:1 instead of the normal 1:4 or 1:6. *Giant cells* that are considerably larger than their neighbors may be formed and possess either one enormous nucleus or several nuclei. *Anaplastic nuclei are variable and bizarre in size and shape.* The chromatin is coarse and clumped, and nucleoli may be of astounding size. More important, *mitoses often are numerous and distinctly atypical*; anarchic multiple spindles may produce tripolar or quadripolar mitotic figures (Fig. 5–5). Also, anaplastic cells usually fail to develop recognizable patterns of orientation to one another (i.e., they lose normal polarity). They may grow in sheets, with total loss of communal structures, such as glands or stratified squamous architecture.

The more differentiated the tumor cell, the more completely it retains the functional capabilities of its normal counterparts. Benign neoplasms and even well-differentiated cancers of endocrine glands frequently elaborate the hormones characteristic of their origin. Well-differentiated squamous cell carcinomas produce keratin (Fig. 5–3), just as well-differentiated hepatocellular carcinomas secrete bile. In other instances, unanticipated functions emerge. Some cancers may elaborate fetal proteins not produced by comparable cells in the adult. Cancers of nonendocrine origin may produce so-called ectopic hormones. For example, certain lung carcinomas may produce adrenocorticotropic hormone (ACTH), parathyroid hormone–like hormone, insulin, glucagon, and others. More is said about these phenomena later. Despite exceptions, *the more rapidly growing and the more anaplastic a tumor, the less likely it is to have specialized functional activity.*

Of relevance in the discussion of differentiation and anaplasia is *dysplasia,* referring to disorderly but non-neoplastic

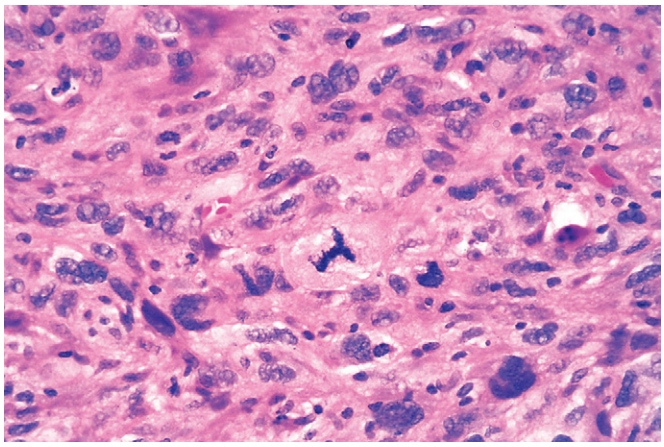


Figure 5–5 High-power detail view of anaplastic tumor cells shows cellular and nuclear variation in size and shape. The prominent cell in the center field has an abnormal tripolar spindle.

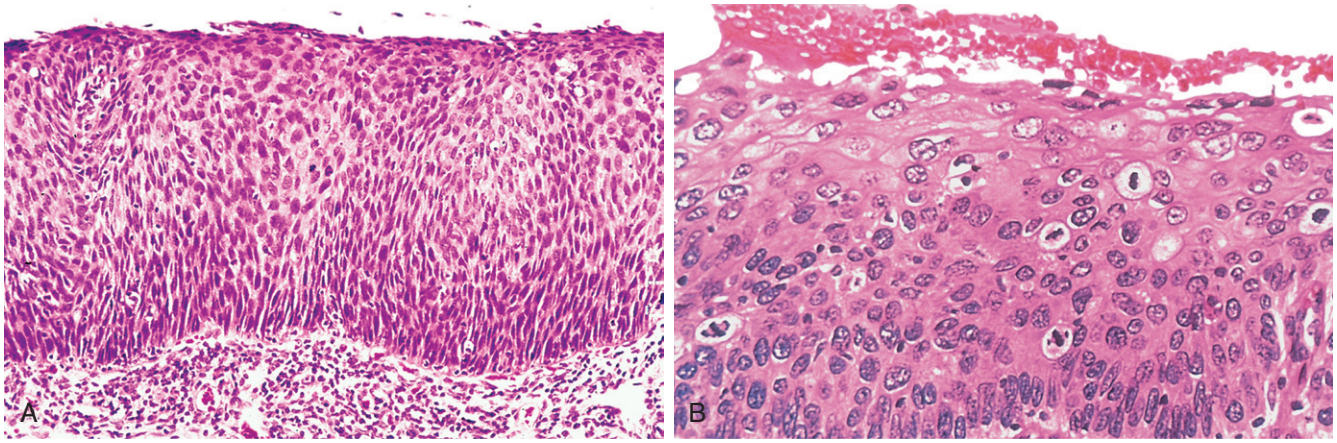


Figure 5–6 Carcinoma in situ. **A,** Low-power view shows that the entire thickness of the epithelium is replaced by atypical dysplastic cells. There is no orderly differentiation of squamous cells. The basement membrane is intact, and there is no tumor in the subepithelial stroma. **B,** High-power view of another region shows failure of normal differentiation, marked nuclear and cellular pleomorphism, and numerous mitotic figures extending toward the surface. The intact basement membrane (*below*) is not seen in this section.

proliferation. Dysplasia is encountered principally in epithelial lesions. It is a *loss in the uniformity of individual cells and in their architectural orientation.* Dysplastic cells exhibit considerable pleomorphism and often possess hyperchromatic nuclei that are abnormally large for the size of the cell. Mitotic figures are more abundant than usual and frequently appear in abnormal locations within the epithelium. In dysplastic stratified squamous epithelium, mitoses are not confined to the basal layers, where they normally occur, but may be seen at all levels and even in surface cells. There is considerable architectural anarchy. For example, the usual progressive maturation of tall cells in the basal layer to flattened squames on the surface may be lost and replaced by a disordered scrambling of dark basal-appearing cells (Fig. 5–6). When dysplastic changes are marked and involve the entire thickness of the epithelium, the lesion is referred to as *carcinoma in situ,* a preinvasive stage of cancer (Chapter 18). Although dysplastic changes often are found adjacent to foci of malignant transformation, and long-term studies of cigarette smokers show that epithelial dysplasia almost invariably antedates the appearance of cancer, *the term dysplasia is not synonymous with cancer; mild to moderate dysplasias that do not involve the entire thickness of the epithelium sometimes regress completely, particularly if inciting causes are removed.*

Rate of Growth

Most benign tumors grow slowly, and most cancers grow much faster, eventually spreading locally and to distant sites (metastasizing) and causing death. There are many exceptions to this generalization, however, and some benign tumors grow more rapidly than some cancers. For example, the rate of growth of leiomyomas (benign smooth muscle tumors) of the uterus is influenced by the circulating levels of estrogens. They may increase rapidly in size during pregnancy and then cease growing, becoming largely fibrocalcific, after menopause. Other influences, such as adequacy of blood supply or pressure constraints, also may affect the growth rate of benign tumors. Adenomas of the pituitary gland locked into the sella turcica have been observed to shrink suddenly. Presumably, they undergo a wave of necrosis as progressive enlargement compresses their blood supply. Despite these caveats and the variation in growth rate from one neoplasm to another, it generally is true that most benign tumors increase in size slowly over the span of months to years.

The rate of growth of malignant tumors usually correlates inversely with their level of differentiation. In other words, poorly differentiated tumors tend to grow more rapidly than do well-differentiated tumors. However, there is wide variation in the rate of growth. Some grow slowly for years and then enter a phase of rapid growth, signifying the emergence of an aggressive subclone of transformed cells. Others grow relatively slowly and steadily; in exceptional instances, growth may come almost to a standstill. Even more exceptionally, some primary tumors (particularly choriocarcinomas) may become totally necrotic, leaving only secondary metastatic implants. Despite these rarities, most cancers progressively enlarge over time, some slowly, others rapidly, but the notion that they "emerge out of the blue" is not true. Many lines of experimental and clinical evidence document that most if not all cancers take years and sometimes decades to evolve into clinically overt lesions. This is true even of "acute" childhood leukemias, which often initiate during fetal development yet manifest as full-blown cancers years later. Rapidly growing malignant tumors often contain central areas of ischemic necrosis, because the tumor blood supply, derived from the host, fails to keep pace with the oxygen needs of the expanding mass of cells.

Cancer Stem Cells and Lineages

The continued growth and maintenance of many tissues that contain short-lived cells, such as the formed elements of the blood and the epithelial cells of the gastrointestinal tract and skin, require a resident population of tissue stem cells that are long-lived and capable of self-renewal. Tissue stem cells are rare and exist in a niche created by support cells, which produce paracrine factors that sustain the stem cells. As described in Chapter 2, tissue stem cells divide asymmetrically to produce two types of daughter

cells—those with limited proliferative potential, which undergo terminal differentiation to form particular tissues, and those that retain stem cell potential. Cancers are immortal and have limitless proliferative capacity, indicating that like normal tissues, they also must contain cells with "stemlike" properties.

The cancer stem cell hypothesis posits that, in analogy with normal tissues, only a special subset of cells within tumors has the capacity for self-renewal. The concept of cancer stem cells has several important implications. Most notably, if cancer stem cells are essential for tumor persistence, it follows that these cells must be eliminated to cure the affected patient. It is hypothesized that, like normal stem cells, cancer stem cells are resistant to conventional therapies, because of their low rate of cell division and the expression of factors, such as multiple drug resistance-1 (MDR-1), that counteract the effects of chemotherapeutic drugs. Thus, the limited success of current therapies could be explained by their failure to kill the malignant stem cells that lie at the root of cancer. Cancer stem cells could arise from normal tissue stem cells or from more differentiated cells that, as part of the transformation process, acquire the property of self-renewal. Studies of certain leukemias (Chapter 11) suggest that both possibilities occur, in that chronic myelogenous leukemia originates from the malignant counterpart of a normal hematopoietic stem cell, whereas certain acute myeloid (myelogenous) leukemias are derived from more differentiated myeloid precursors that acquire an abnormal capacity for self-renewal. The identification of "leukemia stem cells" has spurred the search for cancer stem cells in solid tumors.

Local Invasion

A benign neoplasm remains localized at its site of origin. It does not have the capacity to infiltrate, invade, or metastasize to distant sites, as do malignant neoplasms. For example, as adenomas slowly expand, most develop an enclosing fibrous capsule that separates them from the host tissue. This capsule probably is derived from the stroma of the host tissue as the parenchymal cells atrophy under the pressure of the expanding tumor. The stroma of the tumor itself also may contribute to the capsule (Figs. 5–7 and 5–8). Of note, however, *not all benign neoplasms are encapsulated.* For example, the leiomyoma of the uterus is discretely demarcated from the surrounding smooth muscle by a zone of compressed and attenuated normal myometrium, but there is no well-developed capsule. Nonetheless, a well-defined cleavage plane exists around these lesions. A few benign tumors are neither encapsulated nor discretely defined; such lack of demarcation is particularly likely to be seen in some benign vascular neoplasms of the dermis. These exceptions are pointed out only to emphasize that although encapsulation is the rule in benign tumors, the lack of a capsule does not mean that a tumor is malignant.

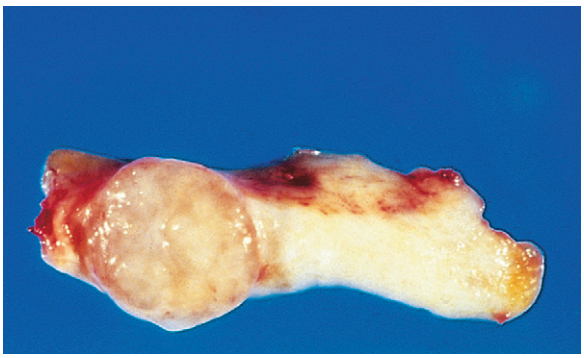


Figure 5–7 Fibroadenoma of the breast. The tan-colored, encapsulated small tumor is sharply demarcated from the whiter breast tissue.

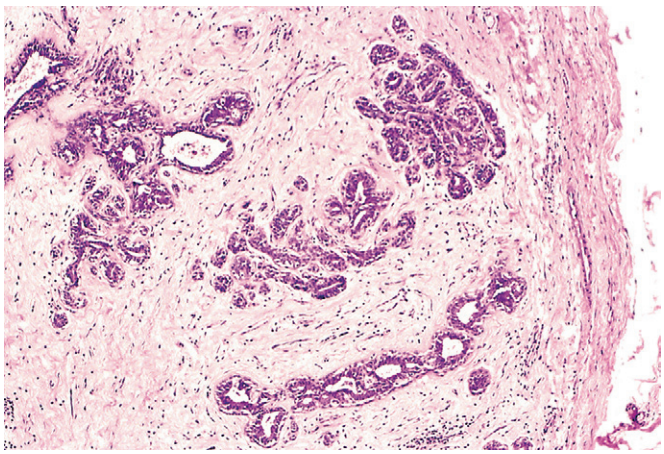


Figure 5–8 Microscopic view of fibroadenoma of the breast seen in Figure 5–7. The fibrous capsule (*right*) sharply delimits the tumor from the surrounding tissue.
(Courtesy of Dr. Trace Worrell, Department of Pathology, University of Texas Southwestern Medical School, Dallas, Texas.)

Cancers grow by progressive infiltration, invasion, destruction, and penetration of the surrounding tissue (Figs. 5–9 and 5–10). They do not develop well-defined capsules. There are, however, occasional instances in which a slowly growing malignant tumor deceptively appears to be encased by the stroma of the surrounding host tissue, but microscopic examination usually reveals tiny crablike feet penetrating the margin and infiltrating adjacent structures. The infiltrative mode of growth makes it necessary to remove a wide margin of surrounding normal tissue when surgical excision of a malignant tumor is attempted.

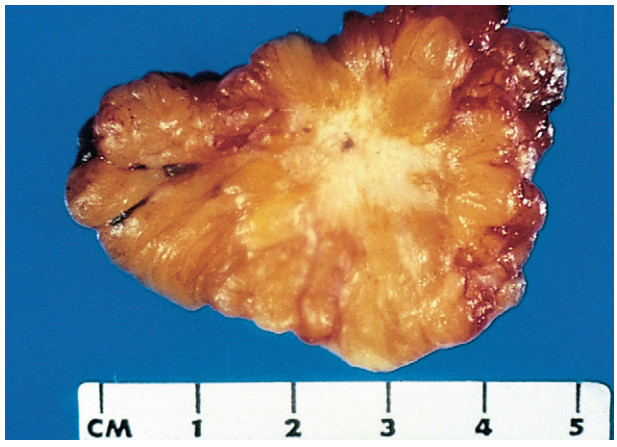


Figure 5–9 Cut section of invasive ductal carcinoma of the breast. The lesion is retracted, infiltrating the surrounding breast substance, and was stony-hard on palpation.

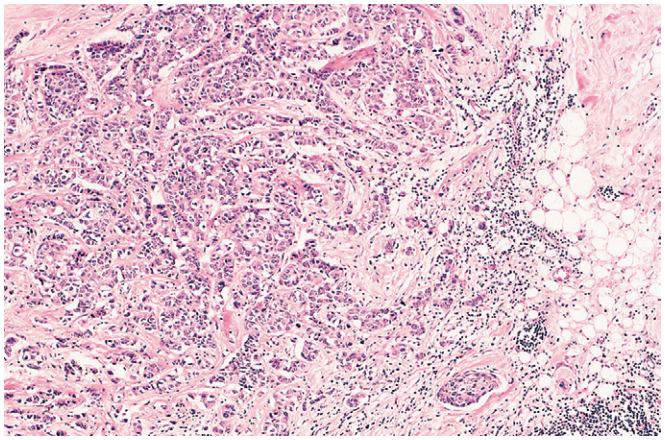


Figure 5–10 Microscopic view of breast carcinoma seen in Figure 5–9 illustrates the invasion of breast stroma and fat by nests and cords of tumor cells (compare with Fig. 5–8). Note the absence of a well-defined capsule.

(Courtesy of Dr. Trace Worrell, Department of Pathology, University of Texas Southwestern Medical School, Dallas, Texas.)

Surgical pathologists carefully examine the margins of resected tumors to ensure that they are devoid of cancer cells (*clean margins*). *Next to the development of metastases, local invasiveness is the most reliable feature that distinguishes malignant from benign tumors.*

Metastasis

Metastases are secondary implants of a tumor that are discontinuous with the primary tumor and located in remote tissues (Fig. 5–11). *More than any other attribute, the property of metastasis identifies a neoplasm as malignant.* Not all cancers have equivalent ability to metastasize, however. At one extreme are basal cell carcinomas of the skin and most primary tumors of the central nervous system, which are highly invasive locally but rarely metastasize. At the other extreme are osteogenic (bone) sarcomas, which usually have metastasized to the lungs at the time of initial discovery.

Approximately 30% of patients with newly diagnosed solid tumors (excluding skin cancers other than melanomas) present with clinically evident metastases. An additional 20% have occult (hidden) metastases at the time of diagnosis.

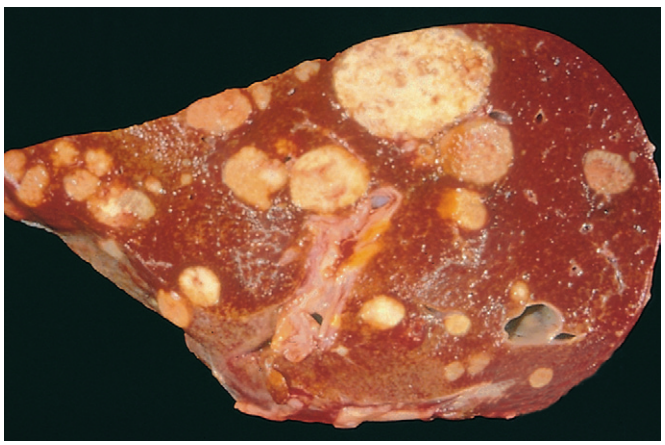


Figure 5–11 A liver studded with metastatic cancer.

In general, the more anaplastic and the larger the primary neoplasm, the more likely is metastatic spread, but as with most rules, there are exceptions. Extremely small cancers have been known to metastasize; conversely, some large and ominous-looking lesions may not. Dissemination strongly prejudices, and may preclude, the possibility of curing the disease, so obviously, short of prevention of cancer, no achievement would confer greater benefit on patients than the prevention of metastases.

Malignant neoplasms disseminate by one of three pathways: (1) seeding within body cavities, (2) lymphatic spread, or (3) hematogenous spread. *Spread by seeding* occurs when neoplasms invade a natural body cavity. This mode of dissemination is particularly characteristic of cancers of the ovary, which often cover the peritoneal surfaces widely. The implants literally may glaze all peritoneal surfaces and yet not invade the underlying tissues. Here is an instance of the ability to reimplant elsewhere that seems to be separable from the capacity to invade. Neoplasms of the central nervous system, such as a medulloblastoma or ependymoma, may penetrate the cerebral ventricles and be carried by the cerebrospinal fluid to reimplant on the meningeal surfaces, either within the brain or in the spinal cord.

Lymphatic spread is more typical of carcinomas, whereas *hematogenous spread* is favored by sarcomas. There are numerous interconnections, however, between the lymphatic and vascular systems, so all forms of cancer may disseminate through either or both systems. The pattern of lymph node involvement depends principally on the site of the primary neoplasm and the natural pathways of local lymphatic drainage. Lung carcinomas arising in the respiratory passages metastasize first to the regional bronchial lymph nodes and then to the tracheobronchial and hilar nodes. Carcinoma of the breast usually arises in the upper outer quadrant and first spreads to the axillary nodes. However, medial breast lesions may drain through the chest wall to the nodes along the internal mammary artery. Thereafter, in both instances, the supraclavicular and infraclavicular nodes may be seeded. In some cases, the cancer cells seem to traverse the lymphatic channels within the immediately proximate nodes to be trapped in subsequent lymph nodes, producing so-called *skip metastases.* The cells may traverse all of the lymph nodes ultimately to reach the vascular compartment by way of the thoracic duct.

A "sentinel lymph node" is the first regional lymph node that receives lymph flow from a primary tumor. It can be identified by injection of blue dyes or radiolabeled tracers near the primary tumor. Biopsy of sentinel lymph nodes allows determination of the extent of spread of tumor and can be used to plan treatment.

Of note, although enlargement of nodes near a primary neoplasm should arouse concern for metastatic spread, it does not always imply cancerous involvement. The necrotic products of the neoplasm and tumor antigens often evoke immunologic responses in the nodes, such as hyperplasia of the follicles (lymphadenitis) and proliferation of macrophages in the subcapsular sinuses (sinus histiocytosis). Thus, histopathologic verification of tumor within an enlarged lymph node is required.

Hematogenous spread is the favored pathway for sarcomas, but carcinomas use it as well. As might be expected,

arteries are penetrated less readily than are veins. With venous invasion, the blood-borne cells follow the venous flow draining the site of the neoplasm, with tumor cells often stopping in the first capillary bed they encounter. Since all portal area drainage flows to the liver, and all caval blood flows to the lungs, *the liver and lungs are the most frequently involved secondary sites in hematogenous dissemination.* Cancers arising near the vertebral column often embolize through the paravertebral plexus; this pathway probably is involved in the frequent vertebral metastases of carcinomas of the thyroid and prostate.

Certain carcinomas have a propensity to grow within veins. Renal cell carcinoma often invades the renal vein to grow in a snakelike fashion up the inferior vena cava, sometimes reaching the right side of the heart. Hepatocellular carcinomas often penetrate portal and hepatic radicles to grow within them into the main venous channels. Remarkably, such intravenous growth may not be accompanied by widespread dissemination.

Many observations suggest that the anatomic localization of a neoplasm and its venous drainage cannot wholly explain the systemic distributions of metastases. For example, prostatic carcinoma preferentially spreads to bone, bronchogenic carcinomas tend to involve the adrenals and the brain, and neuroblastomas spread to the liver and bones. Conversely, skeletal muscles, although rich in capillaries, are rarely the site of secondary deposits. The molecular basis of such tissue-specific homing of tumor cells is discussed later on.

Thus, numerous features of tumors (Fig. 5–12), usually permit the differentiation of benign and malignant neoplasms.

SUMMARY

Characteristics of Benign and Malignant Tumors

- Benign and malignant tumors can be distinguished from one another based on the degree of differentiation, rate of growth, local invasiveness, and distant spread.
- Benign tumors resemble the tissue of origin and are well differentiated; malignant tumors are poorly or completely undifferentiated (anaplastic).
- Benign tumors are slow-growing, whereas malignant tumors generally grow faster.
- Benign tumors are well circumscribed and have a capsule; malignant tumors are poorly circumscribed and invade the surrounding normal tissues.
- Benign tumors remain localized to the site of origin, whereas malignant tumors are locally invasive and metastasize to distant sites.

EPIDEMIOLOGY

Because cancer is a disorder of cell growth and behavior, its ultimate cause must be defined at the cellular and molecular levels. Cancer epidemiology can contribute substantially to knowledge about the origin of cancer. The now well-established concept that cigarette smoking is causally associated with lung cancer arose primarily from epidemiologic studies. A comparison of the incidence rates for colon cancer and dietary patterns in the Western world and in Africa led to the recognition that dietary fat and fiber content may figure importantly in the causation of this

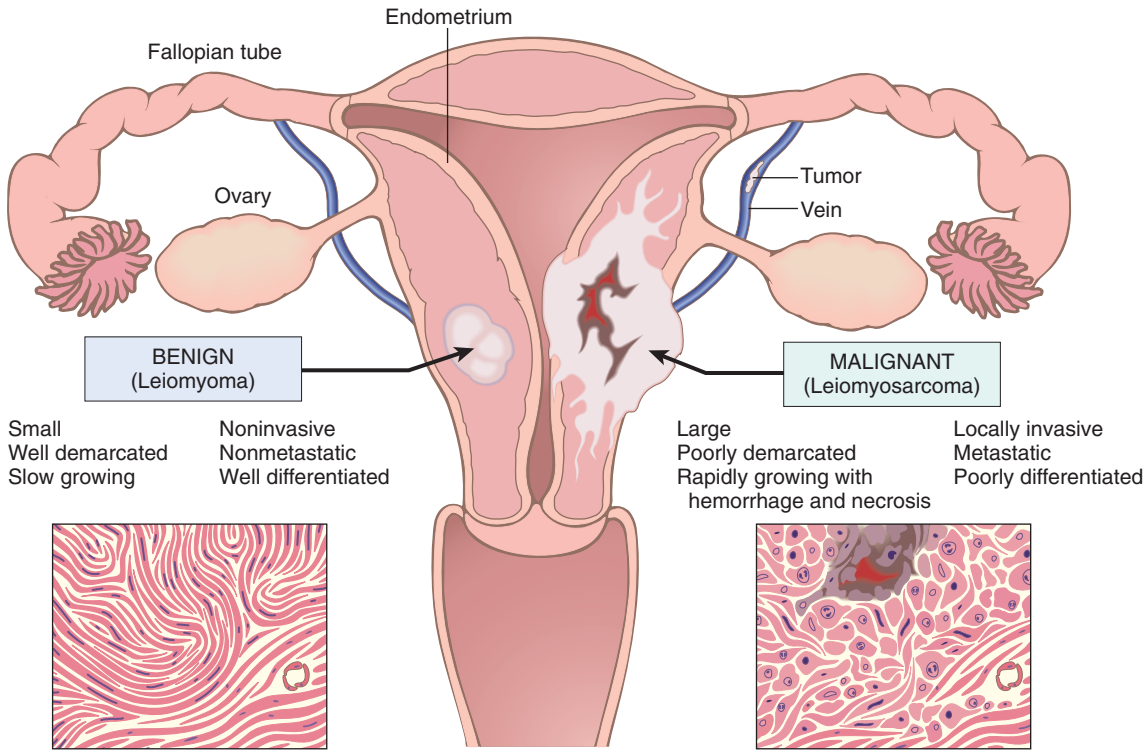


Figure 5–12 Comparison between a benign tumor of the myometrium (leiomyoma) and a malignant tumor of similar origin (leiomyosarcoma).

cancer. Major insights into the causes of cancer can be obtained by epidemiologic studies that relate particular environmental, racial (possibly hereditary), and cultural influences to the occurrence of specific neoplasms. Certain diseases associated with an increased risk of developing cancer (preneoplastic disorders) also provide clues to the pathogenesis of cancer.

The following discussion first summarizes the overall incidence of cancer to provide insight into the magnitude of the cancer problem and then reviews some issues relating to the patient and environment that influence the predisposition to cancer.

Cancer Incidence

Some perspective on the likelihood of developing a specific form of cancer can be gained from national incidence and mortality data. Overall, it is estimated that about *1.5 million* new cancer cases occurred in 2011, and 569,000 people died of cancer in the United States that year. Incidence data for the most common forms of cancer, with the major killers identified, are presented in Figure 5–13.

Over several decades, the death rates for many forms of cancer have changed. Particularly notable is the significant increase in the overall cancer death rate among men that was attributable largely to lung cancer, but this has finally begun to drop. By contrast, the overall death rate among women has fallen slightly, mostly as a result of the decline in death rates for cancers of the uterine cervix, stomach, and large bowel. These welcome trends have more than counterbalanced the striking climb in the rate of lung cancer in women, which not long ago was a relatively uncommon form of neoplasia in this sex. The declining death rate from cervical cancer is directly related to widespread use of cytologic smear studies for early detection of this tumor and its precursor lesions. The development of the human papillomavirus (HPV) vaccine may eliminate this cancer altogether in the coming years. The causes of decline in death rates for cancers of the stomach are obscure; however, there have been speculations about decreasing exposure to dietary carcinogens.

Geographic and Environmental Variables

Although many impressive advances in understanding the molecular pathogenesis of cancer have been made by analyzing hereditary cancers, it is fair to state that environmental factors are the predominant cause of the most common sporadic cancers. This notion is supported by the geographic differences in death rates from specific forms of cancer. For example, death rates from breast cancer are about four to five times higher in the United States and Europe than in Japan. Conversely, the death rate for stomach carcinoma in men and women is about seven times higher in Japan than in the United States. Liver cell carcinoma is relatively infrequent in the United States but is the most lethal cancer among many African populations. Nearly all the evidence indicates that these geographic differences are environmental rather than genetic in origin. Nisei (second-generation Japanese living in the United States) have mortality rates for certain forms of cancer that are intermediate between those in natives of Japan and in Americans who have lived in the United States for many generations. The two rates come closer with each passing generation.

There is no paucity of environmental carcinogens. They lurk in the ambient environment, in the workplace, in food, and in personal practices. They can be as universal as sunlight, can be found particularly in urban settings (e.g., asbestos), or can be limited to a certain occupation (Table 5–2). Certain features of diet have been implicated as possible predisposing influences. Among the possible

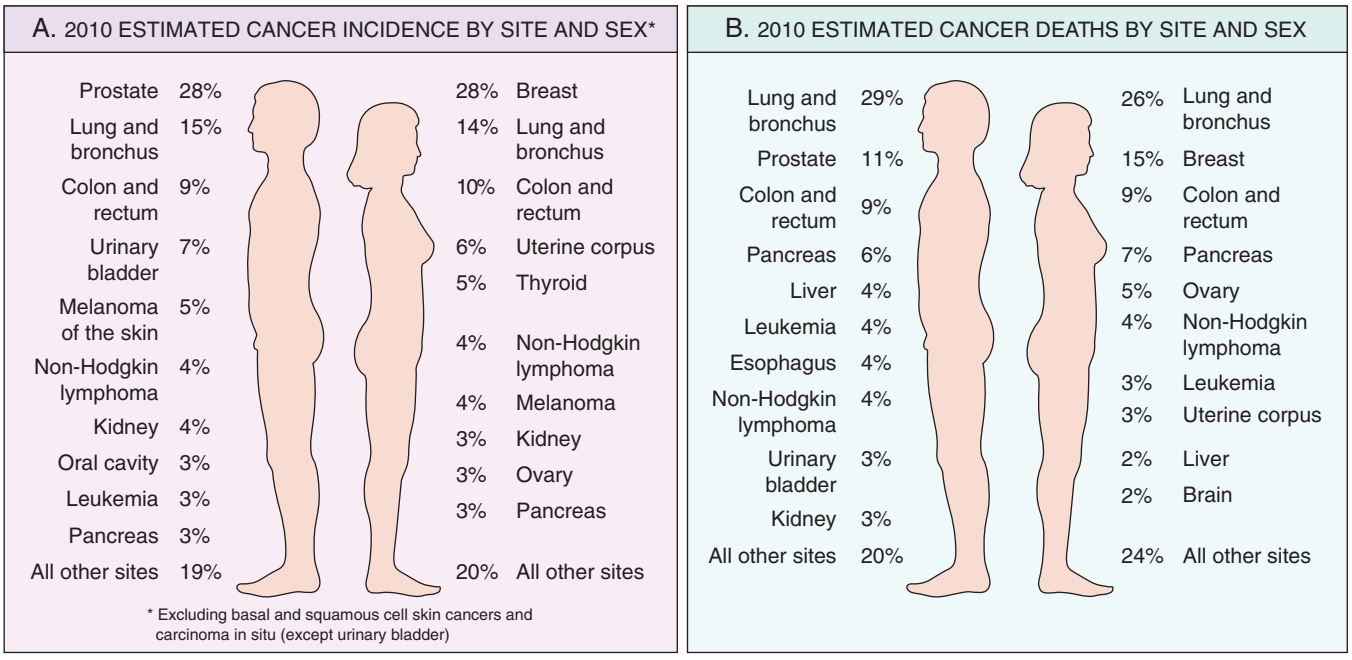


Figure 5–13 Cancer incidence and mortality by site and sex.
(Adapted from Jemal A, Siegel R, Xu J, Ward E: Cancer statistics, 2010. CA Cancer J Clin 60:277–300, 2010.)

Table 5–2 Occupational Cancers

Agent/Group of Agents	Human Cancer Site and Type for Which Reasonable Evidence Is Available	Typical Use/Occurrence
Arsenic and arsenic compounds	Lung, skin, hemangiosarcoma	Byproduct of metal smelting Component of alloys, electrical and semiconductor devices, medications and herbicides, fungicides, and animal dips
Asbestos	Lung, mesothelioma; gastrointestinal tract (esophagus, stomach, large intestine)	Formerly used for many applications because of fire, heat, and friction resistance; still found in existing construction as well as fire-resistant textiles, friction materials (e.g., brake linings), underlayment and roofing papers, and floor tiles
Benzene	Leukemia	Principal component of light oil Many applications exist in printing and lithography, paint, rubber, dry cleaning, adhesives and coatings, and detergents Formerly widely used as solvent and fumigant
Beryllium and beryllium compounds	Lung	Missile fuel and space vehicles Hardener for lightweight compounds metal alloys, particularly in aerospace applications and nuclear reactors
Cadmium and cadmium compounds	Prostate	Uses include yellow pigments and phosphors Found in solders Used in batteries and as alloy and in metal platings and coatings
Chromium compounds	Lung	Component of metal alloys, paints, pigments, and preservatives
Ethylene oxide	Leukemia	Ripening agent for fruits and nuts Used in rocket propellant and chemical synthesis, in fumigants for foodstuffs and textiles, and in sterilants for hospital equipment
Nickel compounds	Nose, lung	Nickel plating Component of ferrous alloys, ceramics, and batteries Byproduct of stainless steel arc welding
Radon and its decay products	Lung	From decay of minerals containing uranium Can be serious hazard in quarries and mines
Vinyl chloride	Angiosarcoma, liver	Refrigerant Monomer for vinyl polymers Adhesive for plastics Formerly used as inert aerosol propellant in pressurized containers

Modified from Stellman JM, Stellman SD: Cancer and workplace. CA Cancer J Clin 46:70–92, 1996, with permission from Lippincott Williams & Wilkins.

environmental influences, the most distressing in terms of prevention are those incurred in personal practices, notably cigarette smoking and chronic alcohol consumption. The risk of cervical cancer is linked to age at first intercourse and the number of sex partners (pointing to a causal role for venereal transmission of the oncogenic virus HPV). There is no escape: It seems that everything people do to earn a livelihood, to subsist, or to enjoy life turns out to be illegal, immoral, or fattening, or—most disturbing—possibly carcinogenic.

Age

In general, the frequency of cancer increases with age. Most cancer deaths occur between ages 55 and 75; the rate declines, along with the population base, after age 75. The rising incidence with age may be explained by the accumulation of somatic mutations associated with the emergence of malignant neoplasms (discussed later). The decline in immune competence that accompanies aging also may be a factor.

Cancer causes slightly more than 10% of all deaths among children younger than 15 years (Chapter 5). The major lethal cancers in children are leukemias, tumors of the central nervous system, lymphomas, and soft tissue and bone sarcomas. As discussed later, study of several childhood tumors, such as retinoblastoma, has provided fundamental insights into the pathogenesis of malignant transformation.

Heredity

The evidence now indicates that for many types of cancer, including the most common forms, there exist not only environmental influences but also hereditary predispositions. Hereditary forms of cancer can be divided into three categories based on their pattern of inheritance (Table 5–3).

Autosomal Dominant Cancer Syndromes

Autosomal dominant cancer syndromes include several well-defined cancers in which inheritance of a single mutant gene greatly increases the risk of developing a tumor. The predisposition to these tumors shows an autosomal dominant pattern of inheritance. Childhood retinoblastoma is the most striking example of this category. Approximately 40% of retinoblastomas are familial. As is discussed later, inherited disabling mutations in a *tumor suppressor gene* are responsible for the development of this tumor in families. Carriers of this gene have a 10,000-fold increased risk of developing retinoblastoma. Unlike those

Table 5–3 Inherited Predisposition to Cancer

Gene(s)	Inherited Predisposition
Autosomal Dominant Cancer Syndromes	
RB	Retinoblastoma
TP53	Li-Fraumeni syndrome (various tumors)
p16INK4A	Melanoma
APC	Familial adenomatous polyposis/colon cancer
NF1, NF2	Neurofibromatosis 1 and 2
BRCA1, BRCA2	Breast and ovarian tumors
MEN1, RET	Multiple endocrine neoplasia 1 and 2
MSH2, MLH1, MSH6	Hereditary nonpolyposis colon cancer
PATCH	Nevoid basal cell carcinoma syndrome
Autosomal Recessive Syndromes of Defective DNA Repair	
Xeroderma pigmentosum	
Ataxia-telangiectasia	
Bloom syndrome	
Fanconi anemia	
Familial Cancers of Uncertain Inheritance	
Breast cancer (not linked to *BRCA1* or *BRCA2*)	
Ovarian cancer	
Pancreatic cancer	

with sporadic retinoblastoma, patients with familial retinoblastoma develop bilateral tumors, and they also have a greatly increased risk of developing a second cancer, particularly osteosarcoma.

Tumors within this group often are associated with a specific marker phenotype. There may be multiple benign tumors in the affected tissue, as occurs in familial polyposis of the colon and in multiple endocrine neoplasia (see Table 5–3). Sometimes, there are abnormalities in tissue that are not the target of transformation (e.g., Lisch nodules and café-au-lait spots in neurofibromatosis type 1) (Chapter 22).

Autosomal Recessive Syndromes of Defective DNA Repair

A group of rare autosomal recessive disorders is collectively characterized by chromosomal or DNA instability and high rates of certain cancers. One of the best-studied is xeroderma pigmentosum, in which DNA repair is defective. This and other familial disorders of DNA instability are described later.

Familial Cancers of Uncertain Inheritance

Virtually all the common types of cancers that occur sporadically have been reported to occur in familial forms where the pattern of inheritance is unclear. Examples are carcinomas of colon, breast, ovary, and brain. *Features that characterize familial cancers include early age at onset, tumors arising in two or more close relatives of the index case, and sometimes multiple or bilateral tumors.* Familial cancers are not associated with specific marker phenotypes. For example, in contrast with the familial adenomatous polyposis syndrome, familial colonic cancers do not arise in preexisting benign polyps. In general, siblings have a relative risk between 2 and 3. Segregation analysis of large families usually reveals that predisposition to the tumors is dominant, but incomplete penetrance or multifactorial inheritance cannot be easily ruled out.

In summary, no more than 5% to 10% of all human cancers fall into one of the three aforementioned categories. What can be said about the influence of heredity in the large preponderance of malignant tumors? There is emerging evidence that the influence of hereditary factors is subtle and sometimes indirect. The genotype may influence the likelihood of developing environmentally induced cancers. For example, polymorphisms in drug-metabolizing enzymes confer genetic predisposition to lung cancer in people who smoke cigarettes. More strikingly, genome-wide association studies (GWAS) in lung cancer, which sought to identify common genetic variants that increase risk for developing cancer, identified variants in a nicotinic acid receptor as being associated with development of lung cancer. Of interest, these variants were strongly associated with the number of cigarettes smoked, suggesting that they indirectly increase lung cancer risk by enhancing the addictiveness of cigarettes.

Acquired Preneoplastic Lesions

Just as some hereditary conditions increase the risk of getting certain cancers, so do certain acquired conditions. These are loosely referred to as *preneoplastic lesions* or simply "*precancers*." These designations are unfortunate because they imply inevitability, but in fact, although such lesions increase the likelihood of malignancy, most do not progress to cancer. In many instances, precursor lesions arise in the setting of chronic tissue injury or inflammation, which may increase the likelihood of malignancy by stimulating continuing regenerative proliferation or by exposing cells to byproducts of inflammation, both of which can lead to somatic mutations (discussed later). Indeed, molecular analyses have shown that many precursor lesions possess some of the genetic lesions found in their associated cancers. Clinically, these precursor lesions are important to recognize, because their removal or reversal may prevent the development of a cancer. A brief listing of some of the chief precursor lesions follows:

- Squamous metaplasia and dysplasia of the bronchial mucosa, seen in habitual smokers—a risk factor for lung cancer
- Endometrial hyperplasia and dysplasia, seen in women with unopposed estrogenic stimulation—a risk factor for endometrial carcinoma
- Leukoplakia of the oral cavity, vulva, or penis, which may progress to squamous cell carcinoma
- Villous adenomas of the colon, associated with a high risk of transformation to colorectal carcinoma

In this context it may be asked, "What is the risk of malignant change in a benign neoplasm?"—or, stated differently, "Are benign tumors precancerous?" In general the answer is no, but inevitably there are exceptions, and perhaps it is better to say that each type of benign tumor is associated with a particular level of risk, ranging from high to virtually nonexistent. For example, adenomas of the colon as they enlarge can undergo malignant transformation in 50% of cases; by contrast, malignant change is extremely rare in leiomyomas of the uterus.

SUMMARY

Epidemiology of Cancer

- The incidence of cancer varies with age, race, geographic factors, and genetic backgrounds. Cancers are most common at the two extremes of age. The geographic variation results mostly from different environmental exposures.
- Most cancers are sporadic, but some are familial. Predisposition to hereditary cancers may be autosomal dominant or autosomal recessive. The former usually are linked to inheritance of a germ line mutation of cancer suppressor genes, whereas the latter typically are associated with inherited defects in DNA repair.
- Familial cancers tend to be bilateral and arise earlier in life than their sporadic counterparts.
- Some acquired diseases, known as preneoplastic disorders, are known to be associated with an increased risk for development of cancer.

CARCINOGENESIS: THE MOLECULAR BASIS OF CANCER

It could be argued that the proliferation of literature on the molecular basis of cancer has outpaced the growth of even the most malignant of tumors. Researchers and students alike can easily get lost in the growing forest of information. Accordingly, a review of some fundamental principles is presented as background for more detailed consideration of the genetic basis of cancer.

- As already discussed, *nonlethal genetic damage lies at the heart of carcinogenesis.* Such genetic damage (or mutation) may be acquired by the action of environmental agents, such as chemicals, radiation, or viruses, or it may be inherited in the germ line. The genetic hypothesis of cancer implies that a tumor mass results from the clonal expansion of a single progenitor cell that has incurred genetic damage (i.e., tumors are monoclonal). This expectation has been realized in all tumors that have been systematically analyzed by genomic sequencing.
- *Four classes of normal regulatory genes – growth-promoting proto-oncogenes, growth-inhibiting tumor suppressor genes, genes that regulate programmed cell death (i.e., apoptosis), and genes involved in DNA repair* – are the principal targets of genetic damage. Collectively, the genetic alterations in tumor cells confer growth and survival advantages over normal cells, as will be evident from the discussion that follows.
- *Oncogenes* are genes that induce a transformed phenotype when expressed in cells. A major discovery in cancer was the realization that most oncogenes are mutated or over expressed versions of normal cellular genes, which are called *proto-oncogenes.* Most known oncogenes encode transcription factors, growth regulating proteins, or proteins involved in cell survival and cell–cell and cell–matrix interactions. They are considered dominant because mutation of a single allele can lead to cellular transformation.
- *Tumor suppressor genes* are genes that normally prevent uncontrolled growth and, when mutated or lost from a cell, allow the transformed phenotype to develop. Usually both normal alleles of tumor suppressor genes must be damaged for transformation to occur. However, recent work has clearly shown that, in some cases, loss of a single allele of a tumor suppressor gene can promote transformation (haploinsufficiency).
- Tumor suppressor genes are usefully placed into two general groups, "*governors*" and "*guardians.*" "Governors" are classic tumor suppressor genes, such as *RB*, where mutation of the gene leads to transformation by removing an important brake on cellular proliferation. "Guardian" genes are responsible for sensing genomic damage. Some of these genes initiate and choreograph a complex "damage control response." This response leads to the cessation of proliferation or, if the damage is too great to be repaired, the induction of apoptosis. *TP53,* the so-called "guardian of the genome," is a prototypic tumor suppressor gene of this type. Other guardian genes are directly involved in recognizing and repairing specific kinds of DNA damage; these are the genes that are mutated in the autosomal recessive syndromes of DNA repair. Mutation of *TP53* or other sensors of genomic damage does not directly transform cells, as loss of guardian function has no direct effect on cellular proliferation or apoptosis. Instead, loss of the guardian genes permits and accelerates the acquisition of mutations in oncogenes and tumor suppressor genes that can lead to the development of cancer. This increase in mutation rate is often referred to as a *mutator phenotype.*
- Genes that regulate apoptosis and DNA repair may act like proto-oncogenes (loss of one copy is sufficient) or tumor suppressor genes (loss of both copies).

Several types of alterations can affect cancer-causing genes and lead to cellular transformation, as detailed in a subsequent section. Presented next is a discussion of the varied genetic lesions that underlie mutation of genes in cancer.

GENETIC LESIONS IN CANCER

The genetic changes that characterize cancer-associated mutations may be subtle (e.g., point mutations or insertions and deletions) or large enough to produce karyotypic changes. Point mutations can either activate or inactivate the resulting protein products. For example, point mutations in proto-oncogenes, such as *RAS* or *EGFR,* frequently result in overactivity of the protein, usually by altering an internal regulatory amino acid and producing a constitutively active protein. However, point mutations in tumor suppressors, such as those affecting *RB* or *TP53* genes, reduce or disable the function of the encoded protein.

Karyotypic Changes in Tumors

The genetic lesion that activates oncogenes or inactivates tumor suppressor genes may be subtle (as described previously) or large enough to be detected in a karyotype. Some cancers have a virtually normal karyotype, while others are markedly aneuploid, with loss and gain of many entire

chromosomes or chromosomal arms. In certain neoplasms, karyotypic abnormalities are nonrandom and common, or even characteristic of a particular tumor. Specific abnormalities have been identified in most leukemias and lymphomas and in an increasing number of nonhematopoietic tumors. The common types of nonrandom structural abnormalities in tumor cells are (1) balanced translocations, (2) deletions, and (3) cytogenetic manifestations of gene amplification.

Balanced Translocations

Balanced translocations are highly associated with certain malignancies, particularly specific kinds of hematopoietic and mesenchymal neoplasms. Translocations can activate proto-oncogenes in two ways:

- Some translocations result in overexpression of proto-oncogenes by removing them from their normal regulatory elements and placing them under control of an inappropriate, highly active promoter. Two different kinds of B cell lymphoma provide cardinal examples of this mechanism. *In more than 90% of cases of Burkitt lymphoma the cells have a translocation, usually between chromosomes 8 and 14, which leads to overexpression of the MYC gene on chromosome 8 by juxtaposition with immunoglobulin heavy chain gene regulatory elements on chromosome 14. In follicular B cell lymphomas, a reciprocal translocation between chromosomes 14 and 18 leads to overexpression of the anti-apoptotic gene, BCL2, on chromosome 18, also driven by immunoglobulin gene elements.*
- Other oncogenic translocations create fusion genes encoding novel chimeric proteins. Most notable is the Philadelphia (Ph) chromosome in chronic myelogenous leukemia, consisting of a reciprocal and balanced translocation between chromosomes 22 and 9 (Fig. 5–14). As a consequence, the derivative chromosome 22 (the Philadelphia chromosome) appears abbreviated. This cytogenetic change, seen in more than 90% of cases of chronic myelogenous leukemia, is a reliable marker of this disease, and the few Ph chromosome–negative cases show molecular evidence of the *BCR-ABL rearrangement, the crucial consequence of Ph translocation. As discussed later, such changes give rise to the BCR-ABL fusion gene with potent tyrosine kinase activity.*

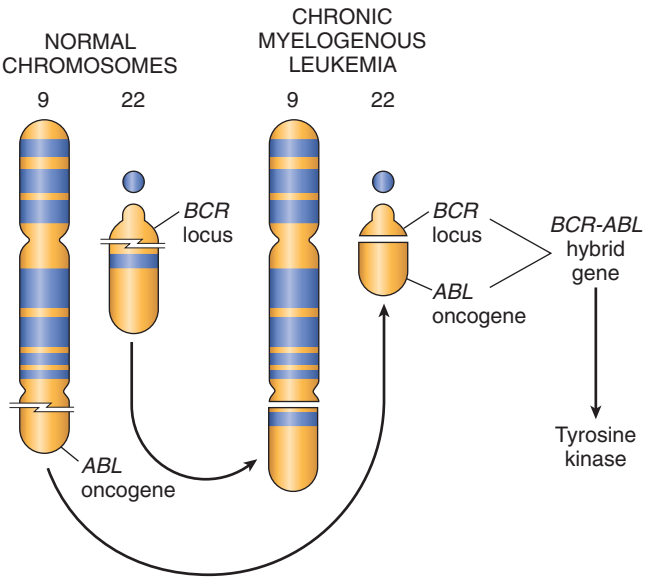


Figure 5–14 The chromosomal translocation and associated oncogene in chronic myelogenous leukemia.

Lymphoid cells are most commonly the targets of gene rearrangements, which may take the form of translocations, inversions, or interstitial deletions, because these cells purposefully make DNA breaks during the processes of antibody or T cell receptor gene recombination. Two other types of mesenchymal tumors, myeloid neoplasms (acute myeloid leukemias and myeloproliferative disorders) and sarcomas, also frequently possess recurrent translocations, such as the t(11;22)(q24;12) translocation in Ewing sarcoma that results in fusion of the EWS transcription factor with Fli-1. The cause of the DNA breaks that lead to chromosomal translocations in myeloid neoplasms and sarcomas is unknown.

Identification of recurrent chromosomal rearrangements in carcinomas has lagged because of the complexity of the karyotypes of these tumors, but novel molecular techniques are beginning to unravel this tangled skein. As with hematologic malignancies and sarcomas, gene rearrangements in solid tumors can contribute to carcinogenesis either by increasing expression of an oncogene or by generation of a novel fusion gene. For example, various *TMPRSS-ETS* fusion genes found in prostate carcinomas place ETS family transcription factor genes under the control of the TMPRSS promoter, which is activated by androgens. The net effect of these rearrangements is the inappropriate, androgen-dependent expression of ETS family transcription factors. Rearrangements of the *HMGA2* gene found in pleomorphic adenomas and other tumors lead to overexpression of the HMGA2 transcription factor through an unusual mechanism; they replace the 3′ untranslated region of *HMGA2* with that of another gene, thus removing key negative regulatory microRNA binding sites. Although the mechanisms are not yet clear, overexpression of HGMA2 or ETS likely promotes carcinogenesis by altering expression of a number of genes that are the targets of these transcription factors. Another uncommon but clinically important type of rearrangement creates an *EML4-ALK* fusion gene, which is present in roughly 4% of lung carcinomas. The EML4-ALK kinase is constitutively active and upregulates signaling through a several progrowth pathways. As discussed later, lung cancers expressing this fusion protein respond to inhibitors of the ALK kinase.

Deletions

Chromosomal deletions are the second most prevalent karyotypic abnormality in tumor cells. Compared with translocations, deletions large enough to be observed karyotypically are more common in nonhematopoietic solid tumors. At a molecular level, however, deletions are commonly found in hematopoietic tumors as well. Deletion of specific regions of chromosomes may result in the loss of particular tumor suppressor genes. Tumor suppressors generally require inactivation of both alleles in order for them to contribute to carcinogenesis. A common mechanism for this is an inactivating point mutation in one allele, followed by deletion of the other, nonmutated allele.

Such deletions result in loss of heterozygosity (LOH), as formerly heterozygous genetic variants will now only have one allele, and all genetic variants within the deleted region will be detected as homozygous. As discussed later, deletions involving 13q14, the site of the *RB* gene, are associated with retinoblastoma, and deletion of 17p is associated with loss of p53.

Gene Amplifications

Proto-oncogenes may be converted to oncogenes by amplification, with consequent overexpression, of otherwise normal proteins. Such amplification may produce several hundred copies of the proto-oncogene in the tumor cell. The amplified genes can be readily detected by molecular hybridization with appropriate DNA probes. In some cases the amplified genes produce chromosomal changes that can be identified microscopically. Two mutually exclusive patterns are seen: multiple small, extrachromosomal structures called "double minutes" and homogeneously staining regions. The latter derive from the insertion of the amplified genes into new chromosomal locations, which may be distant from the normal location of the involved genes; because regions containing amplified genes lack a normal banding pattern, they appear homogeneous in a G-banded karyotype. The most interesting cases of amplification involve *NMYC* in neuroblastoma and *ERBB2* in breast cancers. *NMYC* is amplified in 25% to 30% of neuroblastomas, and the amplification is associated with poor prognosis (Fig. 5–15). *HER2/NEU* (also known as *ERBB2*) amplification occurs in about 20% of breast cancers, and antibody therapy directed against this receptor has proved effective in this subset of tumors.

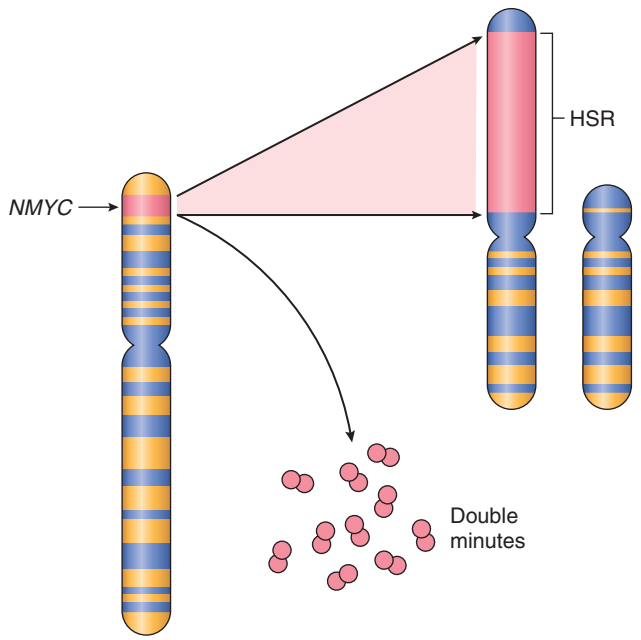


Figure 5–15 Amplification of the *NMYC* gene in human neuroblastoma. The *NMYC* gene, present normally on chromosome 2p, becomes amplified and is seen either as extrachromosomal double minutes or as a chromosomally integrated homogeneous-staining region (HSR). The integration involves other autosomes, such as 4, 9, or 13.

(Modified from Brodeur GM, Seeger RC, Sather H, et al: Clinical implications of oncogene activation in human neuroblastomas. Cancer 58:541, 1986. Reprinted by permission of Wiley-Liss, Inc, a subsidiary of John Wiley & Sons, Inc.)

Aneuploidy

Aneuploidy is defined as a number of chromosomes that is not a multiple of the haploid state; for humans that is a chromosome number that is not a multiple of 23. Aneuploidy is remarkably common in cancers, particularly carcinomas, and was proposed as a cause of carcinogenesis over 100 years ago. Aneuploidy frequently results from errors of the mitotic checkpoint, the major cell cycle control mechanism that acts to prevent chromosome missegregation. The mitotic checkpoint prevents aneuploidy by inhibiting the irreversible transition to anaphase until all of the replicated chromosomes have made productive attachments to spindle microtubules. Complete absence of the mitotic checkpoint leads to rapid cell-autonomous lethality as a consequence of massive chromosome missegregation. However, mechanistic data establishing aneuploidy as a cause of carcinogenesis, rather than a consequence, have been difficult to generate.

MicroRNAs and Cancer

As discussed in Chapter 6, microRNAs (miRNAs) are noncoding, single-stranded RNAs, approximately 22 nucleotides in length, that function as negative regulators of genes. They inhibit gene expression posttranscriptionally by repressing translation or, in some cases, by messenger RNA (mRNA) cleavage. In view of their important function to control cell growth, differentiation, and survival, it is not surprising that accumulating evidence supports a role for miRNAs in carcinogenesis.

As illustrated in Figure 5–16, miRNAs can participate in neoplastic transformation either by increasing the expression of oncogenes or reducing the expression of tumor suppressor genes. If an miRNA inhibits the translation of an oncogene, a reduction in the quantity or function of that miRNA will lead to overproduction of the oncogene product. Conversely, if the target of a miRNA is a tumor suppressor gene, then overactivity of the miRNA can reduce the tumor suppressor protein. Such relationships have already been established by miRNA profiling of several human tumors. For example, downregulation or deletion of certain miRNAs in some leukemias and lymphomas results in increased expression of *BCL2,* the antiapoptotic gene. Thus, by negatively regulating *BCL2,* such miRNAs behave as tumor suppressor genes. Similar miRNA-mediated upregulation of the *RAS* and *MYC* oncogenes also has been detected in lung tumors and in certain B cell leukemias, respectively.

Epigenetic Modifications and Cancer

Epigenetics refers to reversible, heritable changes in gene expression that occur without mutation. Such changes involve posttranslational modifications of histones and DNA methylation, both of which affect gene expression. In normal, differentiated cells, the major portion of the genome is not expressed. These regions of the genome are silenced by DNA methylation and histone modifications. On the other hand, cancer cells are characterized by a global DNA hypomethylation and selective promoter-localized hypermethylation. Indeed, it has become evident

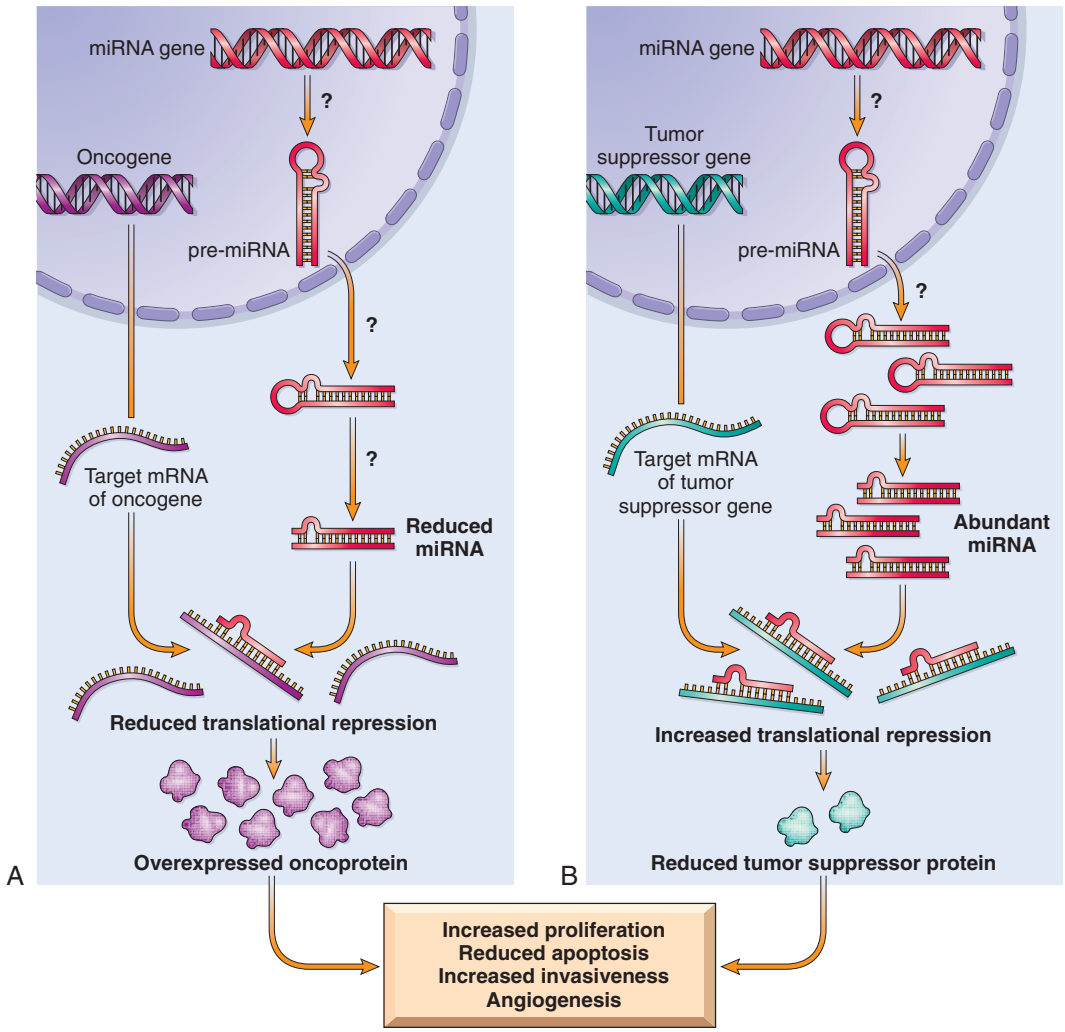


Figure 5–16 Role of microRNAs (miRNAs) in tumorigenesis. **A,** Reduced activity of an miRNA that inhibits translation of an oncogene gives rise to an excess of oncoproteins. **B,** Overactivity of an miRNA that targets a tumor suppression gene reduces the production of the tumor suppressor protein. Question marks in **A** and **B** are meant to indicate that the mechanisms by which changes in the level or activity of miRNA are not entirely known.

during the past several years that tumor suppressor genes are sometimes silenced by hypermethylation of promoter sequences, rather than by mutation. As discussed later, CDKN2A is a complex locus that encodes two tumor suppressors, p14/ARF and p16/INK4a, produced from two different reading frames; p14/ARF is epigenetically silenced in colon and gastric cancers, while p16/INK4a is silenced in a wide variety of cancers. Since this locus produces two tumor suppressors that affect the p53 and Rb pathways, silencing this locus has the pleasing effect (from the cancer's standpoint) of removing two checkpoints with a single alteration. Genome-wide hypomethylation has been shown to cause chromosomal instability and can induce tumors in mice. Thus, epigenetic changes may influence carcinogenesis in many ways. As an added wrinkle, deep sequencing of cancer genomes has identified mutations in genes that regulate epigenetic modifications in a number of cancers. Thus, certain genetic changes in cancers may be selected for because they lead to alterations of the "epigenome" that favor cancer growth and survival.

The epigenetic state of particular cell types—a feature described as the epigenetic context—also dictates their response to signals that control growth and differentiation. As mentioned earlier, epigenetic modifications regulate gene expression, allowing cells with the same genetic make-up (e.g., a neuron and a keratinocyte) to have completely different appearances and functions. In some instances, the epigenetic state of a cell dramatically affects its response to otherwise identical signals. For example, the gene *NOTCH1* has an oncogenic role in T cell leukemia, yet acts as a tumor suppressor in squamous cell carcinomas. As it turns out, activated *NOTCH1* turns on progrowth genes in the epigenetic context of T cell progenitors (e.g., *MYC*) and tumor suppressor genes (e.g., *p21*) in the epigenetic context of keratinocytes.

SUMMARY

Genetic Lesions in Cancer

- Tumor cells may acquire mutations through several means, including point mutations, and nonrandom chromosomal abnormalities that contribute to malignancy; these include

balanced translocations, deletions, and cytogenetic manifestations of gene amplification.

- Balanced translocations contribute to carcinogenesis by overexpression of oncogenes or generation of novel fusion proteins with altered signaling capacity. Deletions frequently affect tumor suppressor genes, whereas gene amplification increases the expression of oncogenes.
- Overexpression of miRNAs can contribute to carcinogenesis by reducing the expression of tumor suppressors, while deletion or loss of expression of miRNAs can lead to overexpression of proto-oncogenes.
- Tumor suppressor genes and DNA repair genes also may be silenced by epigenetic changes, which involve reversible, heritable changes in gene expression that occur not by mutation but by methylation of the promoter.

CARCINOGENESIS: A MULTISTEP PROCESS

Carcinogenesis is a multistep process resulting from the accumulation of multiple genetic alterations that collectively give rise to the transformed phenotype. Many cancers arise from non-neoplastic precursor lesions, which molecular analyses have shown already possess some of the mutations needed to establish a full-blown cancer. Presumably these mutations provide the cells of the precursor lesion with a selective advantage. Once initiated, cancers continue to undergo darwinian selection.

As discussed earlier, malignant neoplasms have several phenotypic attributes, such as excessive growth, local invasiveness, and the ability to form distant metastases. Furthermore, it is well established that over a period of time, many tumors become more aggressive and acquire greater malignant potential. This phenomenon is referred to as *tumor progression* and is not represented simply by an increase in tumor size. Careful clinical and experimental studies reveal that increasing malignancy often is acquired in an incremental fashion. At the molecular level, tumor progression and associated heterogeneity are most likely to result from multiple mutations that accumulate independently in different cells, generating subclones with different characteristics (Fig. 5–17) such as ability to invade, rate of growth, metastatic ability, karyotype, hormonal responsiveness, and susceptibility to antineoplastic drugs. Some of the mutations may be lethal; others may spur cell growth by affecting proto-oncogenes or cancer suppressor genes. *Thus even though most malignant tumors are monoclonal in origin, by the time they become clinically evident their constituent cells may be extremely heterogeneous.*

During progression, tumor cells are subjected to immune and nonimmune selection pressures. For example, cells that are highly antigenic are destroyed by host defenses, whereas those with reduced growth factor requirements are positively selected. A growing tumor, therefore, tends to be enriched for subclones that "beat the odds" and are adept at survival, growth, invasion, and metastasis. Finally, experience has shown that when tumors recur after chemotherapy, the recurrent tumor is almost always resistant to the drug regimen if it is given again. This acquired resistance, too, is a manifestation of selection, as subclones that by chance bear mutations (or perhaps epigenetic alterations) imparting drug resistance survive and are responsible for tumor regrowth. Thus, *genetic evolution and selection can explain two of the most pernicious properties of cancers: the tendency for cancers to become (1) more aggressive and (2) less responsive to therapy over time.*

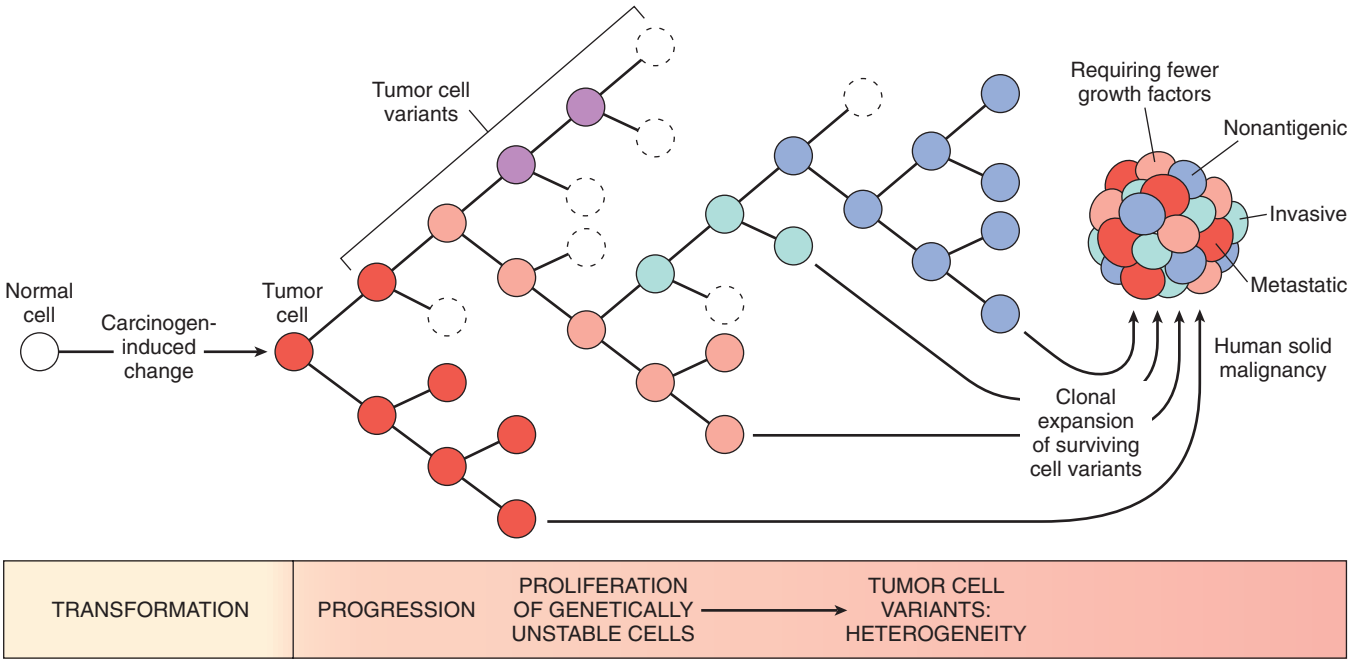


Figure 5–17 Tumor progression and generation of heterogeneity. New subclones arise from the descendants of the original transformed cell by multiple mutations. With progression, the tumor mass becomes enriched for variants that are more adept at evading host defenses and are likely to be more aggressive.

HALLMARKS OF CANCER

This overview serves as background for a more detailed consideration of the molecular pathogenesis of cancer and the carcinogenic agents that inflict genetic damage. In the past 30-some years, hundreds of cancer-associated genes have been discovered. Some, such as *TP53,* are commonly mutated; others, such as *ABL,* are affected only in certain leukemias. Each cancer gene has a specific function, the dysregulation of which contributes to the origin or progression of malignancy. It is best, therefore, to consider cancer-related genes in the context of several fundamental changes in cell physiology, the so-called hallmarks of cancer, which together dictate the malignant phenotype. Six of these are illustrated in Figure 5–18:

- Self-sufficiency in growth signals
- Insensitivity to growth inhibitory signals
- Evasion of cell death
- Limitless replicative potential
- Development of sustained angiogenesis
- Ability to invade and metastasize

To this list may be added two *"emerging" hallmarks* of cancer, reprogramming of energy metabolism and evasion of the immune system, and two *enabling characteristics,* genomic instability and tumor-promoting inflammation.

Mutations in genes that regulate some or all of these cellular traits are seen in every cancer; accordingly, these traits form the basis of the following discussion of the molecular origins of cancer. Of note, by convention, gene symbols are *italicized* but their protein products are not (e.g., *RB* gene and Rb protein, *TP53* and p53, *MYC* and MYC).

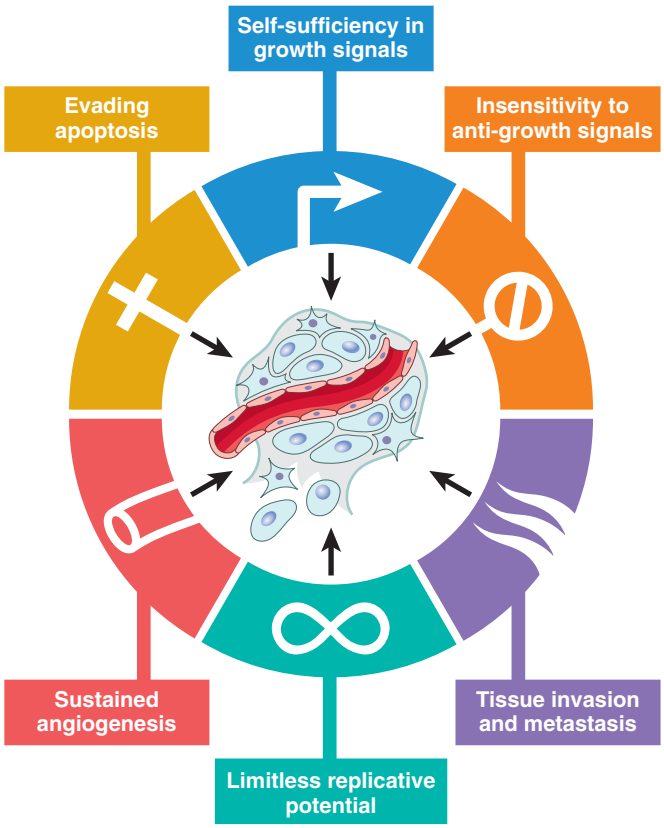


Figure 5–18 Six hallmarks of cancer. Most cancer cells acquire these properties during their development, typically by mutations in the relevant genes.
(From Hanahan D, Weinberg RA: The hallmarks of cancer. Cell 100:57, 2000.)

Self-Sufficiency in Growth Signals

Cancer cells use a number of strategies to drive their proliferation and become insensitive to normal growth regulators. To appreciate these phenomena, it is helpful to review briefly the sequence of events that characterize normal cell proliferation (introduced in Chapter 2). Under physiologic conditions, cell proliferation can be readily resolved into the following steps:

1. The binding of a growth factor to its specific receptor on the cell membrane
2. Transient and limited activation of the growth factor receptor, which in turn activates several signal-transducing proteins on the inner leaflet of the plasma membrane
3. Transmission of the transduced signal across the cytosol to the nucleus by second messengers or a cascade of signal transduction molecules
4. Induction and activation of nuclear regulatory factors that initiate and regulate DNA transcription
5. Entry and progression of the cell into the cell cycle, resulting ultimately in cell division

The mechanisms that endow cancer cells with the ability to proliferate can be grouped according to their role in the growth factor–induced signal transduction cascade and cell cycle regulation. Indeed, each one of the listed steps is susceptible to corruption in cancer cells.

Growth Factors

All normal cells require stimulation by growth factors to undergo proliferation. Most soluble growth factors are made by one cell type and act on a neighboring cell to stimulate proliferation (paracrine action). Normally, cells that produce the growth factor do not express the cognate receptor. This specificity prevents the formation of positive feedback loops within the same cell.

- Many cancer cells acquire growth self-sufficiency by acquiring the ability to synthesize the same growth factors to which they are responsive. For example, many glioblastomas secrete platelet-derived growth factor (PDGF) and express the PDGF receptor, and many sarcomas make both transforming growth factor-α (TGF-α) and its receptor. Similar autocrine loops are fairly common in many types of cancer.
- Another mechanism by which cancer cells acquire growth self-sufficiency is by interaction with stroma. In some cases, tumor cells send signals to activate normal cells in the supporting stroma, which in turn produce growth factors that promote tumor growth.

Growth Factor Receptors and Non-Receptor Tyrosine Kinases

The next group in the sequence of signal transduction is growth factor receptors, and several oncogenes that result from the overexpression or mutation of growth factor

receptors have been identified. Mutant receptor proteins deliver continuous mitogenic signals to cells, even in the absence of the growth factor in the environment. More common than mutations is overexpression of growth factor receptors, which can render cancer cells hyperresponsive to levels of the growth factor that would not normally trigger proliferation. The best-documented examples of overexpression involve the epidermal growth factor (EGF) receptor family. ERBB1, the EGF receptor, is overexpressed in 80% of squamous cell carcinomas of the lung, 50% or more of glioblastomas, and 80% to 100% of epithelial tumors of the head and neck. The gene encoding a related receptor, *HER2/NEU* (*ERBB2*), is amplified in 25% to 30% of breast cancers and adenocarcinomas of the lung, ovary, and salivary glands. These tumors are exquisitely sensitive to the mitogenic effects of small amounts of growth factors, and a high level of HER2/NEU protein in breast cancer cells is a harbinger of poor prognosis. The significance of *HER2/NEU* in the pathogenesis of breast cancers is illustrated dramatically by the clinical benefit derived from blocking the extracellular domain of this receptor with anti-*HER2/NEU* antibodies. Treatment of breast cancer with anti-HER2/NEU antibody is an elegant example of "bench to bedside" medicine.

Downstream Signal-Transducing Proteins

A relatively common mechanism by which cancer cells acquire growth autonomy is mutations in genes that encode various components of the signaling pathways downstream of growth factor receptors. These signaling proteins couple growth factor receptors to their nuclear targets. They receive signals from activated growth factor receptors and transmit them to the nucleus, either through second messengers or through a cascade of phosphorylation and activation of signal transduction molecules. Two important members in this category are *RAS* and *ABL*. Each of these is discussed briefly next.

RAS Protein. *RAS* is the most commonly mutated proto-oncogene in human tumors. Indeed, approximately 30% of all human tumors contain mutated versions of the *RAS* gene, and the frequency is even higher in some specific cancers (e.g., colon and pancreatic adenocarcinomas).

- RAS is a member of a family of small G proteins that bind guanosine nucleotides (guanosine triphosphate [GTP] and guanosine diphosphate [GDP]), similar to the larger trimolecular G proteins.
- Normal RAS proteins flip back and forth between an excited signal-transmitting state and a quiescent state. RAS proteins are inactive when bound to GDP; stimulation of cells by growth factors such as EGF and PDGF leads to exchange of GDP for GTP and subsequent conformational changes that generate active RAS (Fig. 5–19). This excited signal-emitting state is short-lived, however, because the intrinsic guanosine triphosphatase (GTPase) activity of RAS hydrolyzes GTP to GDP, releasing a phosphate group and returning the protein to its quiescent GDP-bound state. The GTPase activity of activated RAS protein is magnified dramatically by a family of GTPase-activating proteins (GAPs), which act as molecular brakes that prevent uncontrolled RAS activation by favoring hydrolysis of GTP to GDP.

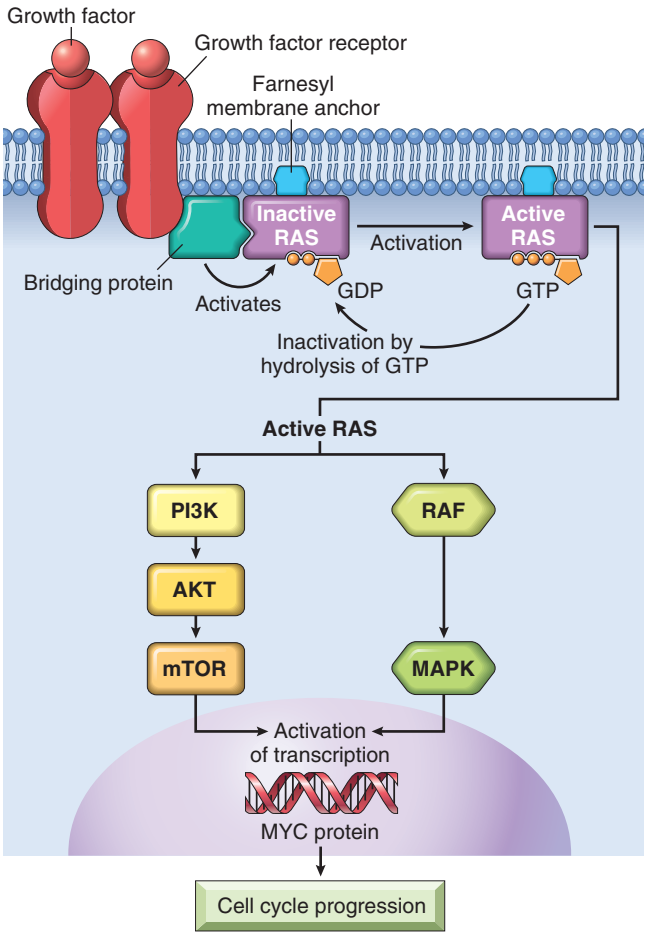


Figure 5–19 Model for action of *RAS* genes. When a normal cell is stimulated through a growth factor receptor, inactive (GDP-bound) RAS is activated to a GTP-bound state. Activated RAS transduces proliferative signals to the nucleus along two pathways: the so-called RAF/ERK/MAP kinase pathway and the PI3 kinase/AKT pathway. GDP, guanosine diphosphate; GTP, guanosine triphosphate; MAP, mitogen-activated protein; PI3, phosphatidylinositol-3.

- The activated RAS stimulates downstream regulators of proliferation by two distinct pathways that converge on the nucleus and flood it with signals for cell proliferation. While details of the signaling cascades (some of which are illustrated in Fig. 5–19) downstream of RAS are not discussed here, an important point is that mutational activation of these "messengers" to the nucleus can mimic the growth promoting effects of activated RAS. For example, BRAF, which lies in the so-called RAF/ERK/MAP kinase pathway, is mutated in more than 60% of melanomas. Mutations of PI3 kinase in the PI3K/AKT pathway also occur with high frequency in some tumor types. Indeed, it appears that activating mutations of RAS as well as its downstream signaling molecules are very common in a wide variety of tumors.

The RAS protein most commonly is activated by point mutations in amino acid residues that are either within the GTP-binding pocket or in the enzymatic region essential for GTP hydrolysis. Both kinds of mutations interfere with GTP hydrolysis, which is essential to inactivate RAS. RAS is thus trapped in its activated, GTP-bound form, and the

cell is forced into a continuously proliferating state. It follows from this scenario that the consequences of mutations in RAS protein would be mimicked by loss-of-function mutations in the GAPs with a failure to simulate GTP hydrolysis and thereby restrain normal RAS proteins. Indeed, disabling mutation of neurofibromin-1 (NF-1), a GAP, is associated with familial neurofibromatosis type 1 (Chapter 22).

ABL. In addition to *RAS,* several non–receptor-associated tyrosine kinases function as signal transduction molecules. In this group, *ABL* is the most well defined with respect to carcinogenesis.

- The *ABL* proto-oncogene has tyrosine kinase activity that is dampened by internal negative regulatory domains. In chronic myelogenous leukemia and certain acute leukemias, a part of the *ABL* gene is translocated from its normal abode on chromosome 9 to chromosome 22, where it fuses with part of the breakpoint cluster region (*BCR*) gene. The BCR-ABL hybrid protein maintains the tyrosine kinase domain; the BCR domain self-associates, a property that unleashes a constitutive tyrosine kinase activity. Of interest, there is cross-talk between BCR-ABL and RAS pathways, since BCR-ABL protein activates all of the signals that are downstream of RAS.
- The crucial role of *BCR-ABL* in transformation has been confirmed by the dramatic clinical response of patients with chronic myelogenous leukemia to BCR-ABL kinase inhibitors. The prototype of this kind of drug, imatinib mesylate (Gleevec), galvanized interest in design of drugs that target specific molecular lesions found in various cancers (so-called *targeted therapy*). BCR-ABL also is an example of the concept of *oncogene addiction,* wherein a tumor is profoundly dependent on a single signaling molecule. BCR-ABL fusion gene formation is an early, perhaps initiating, event that drives leukemogenesis. Development of leukemia probably requires other collaborating mutations, but the transformed cell continues to depend on BCR-ABL for signals that mediate growth and survival. BCR-ABL signaling can be seen as the central lodgepole around which the structure is built. If the lodgepole is removed by inhibition of the BCR-ABL kinase, the structure collapses. In view of this level of dependency, it is not surprising that acquired resistance of tumors to BCR-ABL inhibitors often is due to the outgrowth of a subclone with a mutation in BCR-ABL that prevents binding of the drug to the BCR-ABL protein.

Nuclear Transcription Factors

Ultimately, all signal transduction pathways enter the nucleus and have an impact on a large bank of responder genes that orchestrate the cell's orderly advance through the mitotic cycle. Indeed, the ultimate consequence of signaling through oncoproteins such as RAS or ABL is inappropriate and continuous stimulation of nuclear transcription factors that drive the expression of growth-promoting genes. Growth autonomy may thus be a consequence of mutations affecting genes that regulate transcription of DNA. A host of oncoproteins, including products of the *MYC, MYB, JUN, FOS,* and *REL* oncogenes, function as transcription factors that regulate the expression of growth-promoting genes, such as cyclins. Of these, the *MYC* gene is involved most commonly in human tumors.

The MYC protein can either activate or repress the transcription of other genes. Those activated by MYC include several growth-promoting genes, including cyclin-dependent kinases (CDKs), whose products drive cells into the cell cycle (discussed next). Genes repressed by MYC include the *CDK* inhibitors (*CDKI*s). Thus, dysregulation of MYC promotes tumorigenesis by increasing expression of genes that promote progression through the cell cycle and repressing genes that slow or prevent progression through the cell cycle. MYC also is a key regulator of intermediate metabolism, upregulating genes that promote aerobic glycolysis (the so-called Warburg effect, described later) and the increased utilization of glutamine, two metabolic changes that are hallmarks of cancer cells. Dysregulation of the *MYC* gene resulting from a t(8;14) translocation occurs in Burkitt lymphoma, a B cell tumor. *MYC* also is amplified in breast, colon, lung, and many other cancers; the related *NMYC* and *LMYC* genes are amplified in neuroblastomas and small cell cancers of lung.

Cyclins and Cyclin-Dependent Kinases

The ultimate outcome of all growth-promoting stimuli is the entry of quiescent cells into the cell cycle. Cancers may become autonomous if the genes that drive the cell cycle become dysregulated by mutations or amplification. Before further consideration of this aspect of carcinogenesis, a brief review of the normal cell cycle is warranted (Fig. 5–20).

The Normal Cell Cycle

Cell proliferation is a tightly controlled process that involves a large number of molecules and interrelated pathways. The replication of cells is stimulated by growth factors or by signaling from ECM components through integrins. To achieve DNA replication and division, the cell goes through a tightly controlled sequence of events known as the cell cycle. The cell cycle consists of G_1 (presynthetic), S (DNA synthesis), G_2 (premitotic), and M (mitotic) phases. Quiescent cells that have not entered the cell cycle are in the G_0 state. Each cell cycle phase is dependent on the proper activation and completion of the previous ones and the cycle stops at a place at which an essential gene function is deficient. Because of its central role in maintaining tissue homeostasis and regulating physiologic growth processes such as regeneration and repair, the cell cycle has multiple checkpoints, particularly during emergence from G_0 into G_1 and the transition from G_1 to S phase.

Cells can enter G_1 either from G_0 (quiescent cells) or after completing mitosis (continuously replicating cells). Quiescent cells must first go through the transition from G_0 to G_1, the first decision step, which functions as a gateway to the cell cycle. Cells in G_1 progress through the cell cycle and reach a critical stage at the G_1-S transition, known as a restriction point, a rate-limiting step for replication. On passing this restriction point, normal cells become irreversibly committed to DNA replication. The cell cycle is tightly controlled by activators and inhibitors.

- Progression through the cell cycle, particularly at the G_1-S transition, is regulated by proteins called *cyclins,* so

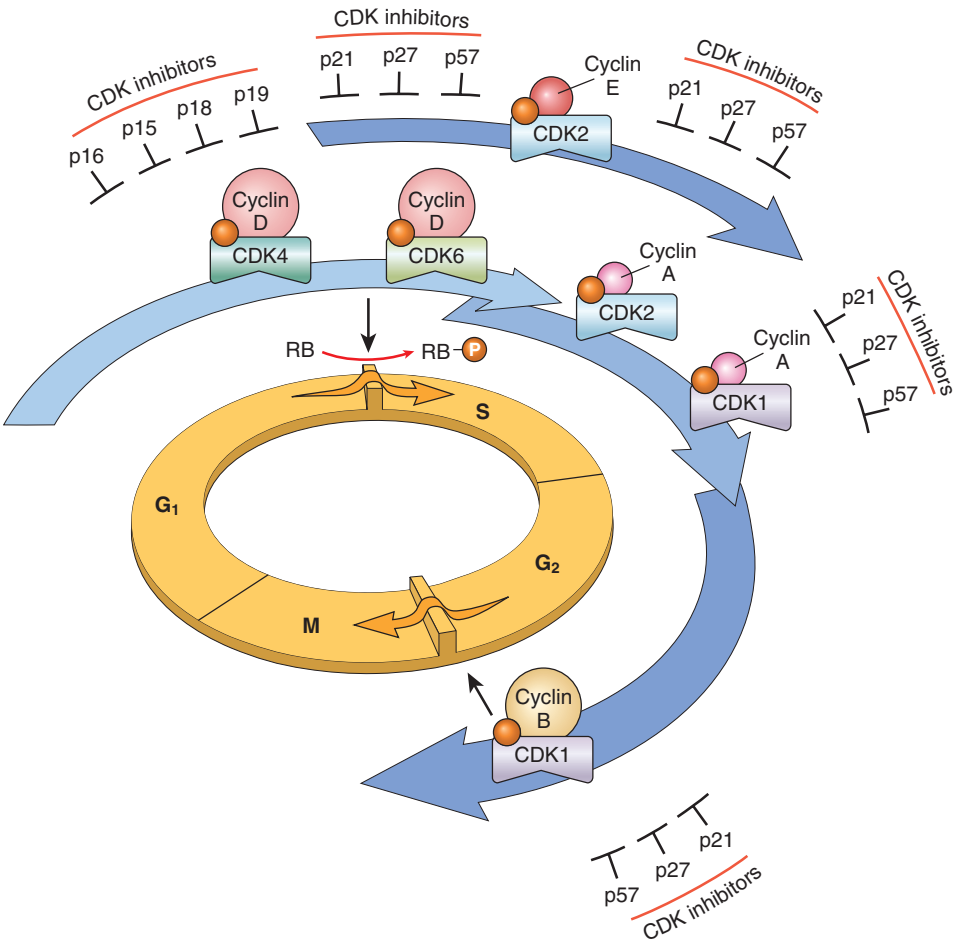


Figure 5–20 Role of cyclins, cyclin-dependent kinases (CDKs), and CDK inhibitors in regulating the cell cycle. The *shaded arrows* represent the phases of the cell cycle during which specific cyclin–CDK complexes are active. As illustrated, cyclin D–CDK4, cyclin D–CDK6, and cyclin E–CDK2 regulate the G_1-to-S transition by phosphorylating the Rb protein (pRb). Cyclin A–CDK2 and cyclin A–CDK1 are active in the S phase. Cyclin B–CDK1 is essential for the G_2-to-M transition. Two families of CDK inhibitors can block activity of CDKs and progression through the cell cycle. The so-called INK4 inhibitors, composed of p16, p15, p18, and p19, act on cyclin D–CDK4 and cyclin D–CDK6. The other family of three inhibitors, p21, p27, and p57, can inhibit all CDKs.

called because of the cyclic nature of their production and degradation, and associated enzymes, the *cyclin-dependent kinases* (CDKs). CDKs acquire catalytic activity by binding to and forming complexes with the cyclins. The orderly progression of cells through the various phases of the cell cycle is orchestrated by CDKs, which are activated by binding to the cyclins.

- The CDK–cyclin complexes phosphorylate crucial target proteins that drive the cell through the cell cycle. On completion of this task, cyclin levels decline rapidly. More than 15 cyclins have been identified; cyclins D, E, A, and B appear sequentially during the cell cycle and bind to one or more CDKs. The cell cycle may thus be seen as a relay race in which each leg is regulated by a distinct set of cyclins: As one set of cyclins leaves the track, the next set takes over (Fig. 5–20). Activated CDKs in these complexes drive the cell cycle by phosphorylating proteins that regulate cell cycle transitions. One such protein is the retinoblastoma protein (Rb), discussed later.
- The activity of CDK–cyclin complexes is regulated by CDK inhibitors (CDKIs), which enforce cell cycle checkpoints. Embedded in the cell cycle are surveillance mechanisms that are geared to sensing damage to DNA and chromosomes. These quality control checks are called *checkpoints;* they ensure that cells with damaged DNA or chromosomes do not complete replication. The G_1-S checkpoint monitors the integrity of DNA before DNA replication, whereas the G_2-M checkpoint checks DNA after replication and monitors whether the cell can safely enter mitosis. When cells sense DNA damage, checkpoint activation delays the cell cycle and triggers DNA repair mechanisms. If DNA damage is too severe to be repaired, the cells are eliminated by apoptosis, or enter a nonreplicative state called senescence, primarily through p53-dependent mechanisms, discussed later on. Mutations in genes regulating these checkpoints allow cells with damaged DNA to divide, producing daughter cells carrying mutations.
- There are several families of CDKIs. One family, composed of three proteins called p21 (CDKN1A), p27 (CDKN1B), and p57 (CDKN1C), inhibits the CDKs broadly, whereas the other family of CDKIs has selective effects on cyclin CDK4 and cyclin CDK6. The four

members of this family—p15 (CDKN2B), p16 (CDKN2A), p18 (CDKN2C), and p19 (CDKN2D)—are sometimes called INK4 (A to D) proteins.

Alterations in Cell Cycle Control Proteins in Cancer Cells

With this background it is easy to appreciate that mutations that dysregulate the activity of cyclins and CDKs would favor cell proliferation. Indeed, all cancers appear to have genetic lesions that disable the G_1-S checkpoint, causing cells to continually reenter the S phase. For unclear reasons, particular lesions vary widely in frequency across tumor types.

- Mishaps increasing the expression of cyclin D or CDK4 seem to be a common event in neoplastic transformation. The cyclin D genes are overexpressed in many cancers, including those affecting the breast, esophagus, liver, and a subset of lymphomas and plasma cell tumors. Amplification of the *CDK4* gene occurs in melanomas, sarcomas, and glioblastomas. Mutations affecting cyclins B and E and other CDKs also occur, but they are much less frequent than those affecting cyclin CDK4.
- The CDKIs frequently are disabled by mutation or gene silencing in many human malignancies. Germline mutations of *CDKN2A* are present in 25% of melanoma-prone kindreds. Somatically acquired deletion or inactivation of *CDKN2A* is seen in 75% of pancreatic carcinomas, 40% to 70% of glioblastomas, 50% of esophageal cancers, and 20% of non–small cell lung carcinomas, soft tissue sarcomas, and bladder cancers.

A final consideration of importance in a discussion of growth-promoting signals is that the increased production of oncoproteins does not by itself lead to sustained proliferation of cancer cells. There are two built-in mechanisms, cell senescence and apoptosis, that oppose oncogene-mediated cell growth. As discussed later, genes that regulate these two braking mechanisms must be disabled to allow unopposed action of oncogenes.

SUMMARY

Oncogenes That Promote Unregulated Proliferation (Self-Sufficiency in Growth Signals)

Proto-oncogenes: normal cellular genes whose products promote cell proliferation

Oncogenes: mutant or overexpressed versions of proto-oncogenes that function autonomously without a requirement for normal growth-promoting signals

Oncoproteins promote uncontrolled cell proliferation by several mechanisms:

- Stimulus-independent expression of growth factor and its receptor, setting up an autocrine loop of cell proliferation
  - PDGF–PDGF receptor in brain tumors
- Mutations in genes encoding growth factor receptors or tyrosine kinases leading to constitutive signaling
  - EGF receptor family members, including HER2/NEU (breast, lung, and other tumors)
  - Fusion of ABL tyrosine kinase with BCR protein in certain leukemias generates a hybrid protein with constitutive kinase activity.
- Mutations in genes encoding signaling molecules
  - RAS commonly is mutated in human cancers and normally flips between resting GDP-bound state and active GTP-bound state; mutations block hydrolysis of GTP to GDP, leading to unchecked signaling.
- Overproduction or unregulated activity of transcription factors
  - Translocation of *MYC* in some lymphomas leads to overexpression and unregulated expression of its target genes controlling cell cycling and survival.
- Mutations that activate cyclin genes or inactivate negative regulators of cyclins and cyclin-dependent kinases
  - Complexes of cyclins with CDKs drive the cell cycle by phosphorylating various substrates. CDKs are controlled by inhibitors; mutations in genes encoding cyclins, CDKs, and CDK inhibitors result in uncontrolled cell cycle progression. Such mutations are found in a wide variety of cancers including melanomas, brain, lung, and pancreatic cancer.

Insensitivity to Growth Inhibitory Signals

Isaac Newton theorized that every action has an equal and opposite reaction. Although Newton was not a cancer biologist, his formulation holds true for cell growth. Whereas oncogenes encode proteins that promote cell growth, the products of tumor suppressor genes apply brakes to cell proliferation. Disruption of such genes renders cells refractory to growth inhibition and mimics the growth-promoting effects of oncogenes. The following discussion describes tumor suppressor genes, their products, and possible mechanisms by which loss of their function contributes to unregulated cell growth.

RB *Gene: Governor of the Cell Cycle*

It is useful to begin with the retinoblastoma gene (*RB*), the first tumor suppressor gene to be discovered and, as it happens, a prototypical representative. As with many advances in medicine, the discovery of tumor suppressor genes was accomplished by the study of a rare disease—in this case, retinoblastoma, an uncommon childhood tumor. Approximately 60% of retinoblastomas are sporadic, and the remaining ones are familial, the predisposition to develop the tumor being transmitted as an autosomal dominant trait. To account for the sporadic and familial occurrence of an identical tumor, Knudson, in 1974, proposed his now famous *two-hit* hypothesis, which in molecular terms can be stated as follows:

- Two mutations (*hits*) are required to produce retinoblastoma. These involve the *RB* gene, which has been mapped to chromosomal locus 13q14. Both of the normal alleles of the *RB* locus must be inactivated (hence the two hits) for the development of retinoblastoma (Fig. 5–21).
- In familial cases, children inherit one defective copy of the *RB* gene in the germ line; the other copy is normal.

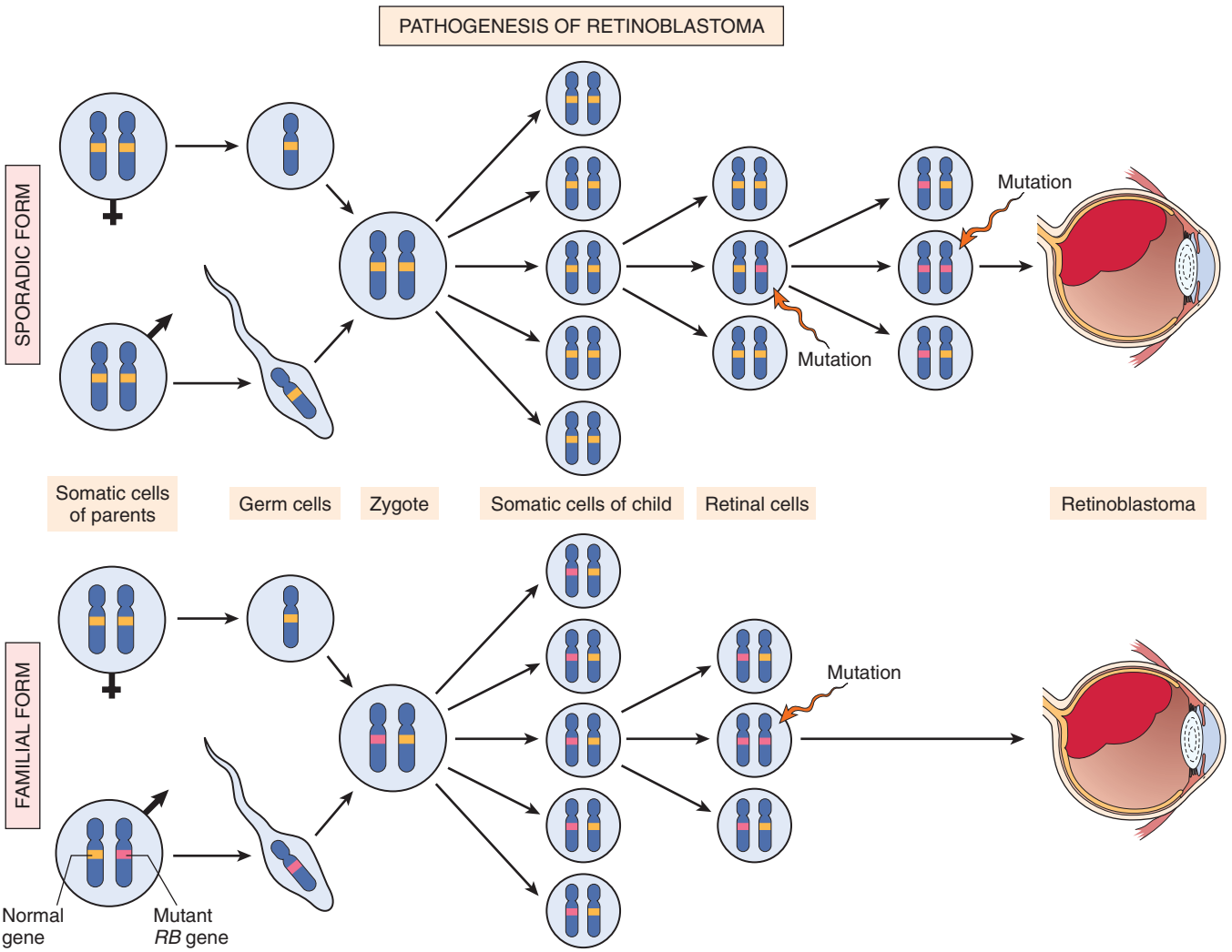


Figure 5–21 Pathogenesis of retinoblastoma. Two mutations of the *RB* chromosomal locus, on 13q14, lead to neoplastic proliferation of the retinal cells. In the familial form, all somatic cells inherit one mutant *RB* gene from a carrier parent. The second mutation affects the *RB* locus in one of the retinal cells after birth. In the sporadic form, both mutations at the *RB* locus are acquired by the retinal cells after birth.

Retinoblastoma develops when the normal *RB* gene is lost in retinoblasts as a result of somatic mutation. Because in retinoblastoma families only a single somatic mutation is required for expression of the disease, the familial transmission follows an autosomal dominant inheritance pattern.

- In sporadic cases, both normal *RB* alleles are lost by somatic mutation in one of the retinoblasts. The end result is the same: a retinal cell that has lost both of the normal copies of the *RB* gene becomes cancerous.

Although the loss of normal *RB* genes initially was discovered in retinoblastomas, it is now evident that homozygous loss of this gene is a fairly common feature of several tumors, including breast cancer, small cell cancer of the lung, and bladder cancer. Patients with familial retinoblastoma also are at greatly increased risk for development of osteosarcomas and some soft tissue sarcomas.

At this point, some clarification of terminology is in order: A cell heterozygous at the *RB* locus is not neoplastic. Tumors develop when the cell loses its normal *RB* gene copy and thus becomes *homozygous* for the mutant allele.

In principle, antigrowth signals can prevent cell proliferation by several complementary mechanisms. The signal may cause dividing cells to enter G_0 (quiescence), where they remain until external cues prod their reentry into the proliferative pool. Alternatively, the cells may enter a postmitotic, differentiated pool and lose replicative potential. Nonreplicative senescence, alluded to earlier, is another escape mechanism from sustained cell growth. And, as a last-ditch effort, the cells may be programmed for death by apoptosis. As we shall see, tumor suppressor genes have all these "tricks" in their toolbox designed to halt wayward cells from becoming malignant.

The subsequent discussion of growth inhibitory mechanisms and their evasion focuses initially on the prototypical tumor suppressor gene, the *RB* gene.

SUMMARY

Insensitivity to Growth Inhibitory Signals

- Tumor suppressor genes encode proteins that inhibit cellular proliferation by regulating the cell cycle. Unlike oncogenes, both copies of the gene must be dysfunctional for tumor development to occur.
- In cases with familial predisposition for development of tumors, affected persons inherit one defective (nonfunctional) copy of a tumor suppressor gene and lose the second one through somatic mutation. In sporadic cases, both copies are lost through somatic mutations.

The *RB* gene product is a DNA-binding protein that is expressed in every cell type examined, where it exists in an *active hypophosphorylated state* and an *inactive hyperphosphorylated state*. The importance of Rb lies in its regulation of the G_1/S checkpoint, the portal through which cells must pass before DNA replication commences.

As background for an understanding of how tumor suppressors function, it is useful to briefly revisit the cell cycle: In embryos, cell divisions proceed at an amazing clip, with DNA replication beginning immediately after mitosis ends. As development proceeds, however, two gaps are incorporated into the cell cycle: gap 1 (G_1) between mitosis (M) and DNA replication (S), and gap 2 (G_2) between DNA replication (S) and mitosis (M) (Fig. 5–20). Although each phase of the cell cycle circuitry is monitored carefully, the transition from G_1 to S is believed to be an extremely important checkpoint in the cell cycle "clock." Once cells cross the G_1 checkpoint they can pause the cell cycle for a time, but they are obligated to complete mitosis. In G_1, however, cells can remove themselves entirely from the cell cycle, either temporarily (quiescence, or G_0) or permanently (senescence). Indeed, during development, as cells become terminally differentiated, they exit the cell cycle and enter G_0. Cells in G_0 remain there until external cues, such as mitogenic signaling, push them back into the cell cycle. In G_1, therefore, diverse signals are integrated to determine whether the cell should progress through the cell cycle, or exit the cell cycle and differentiate, and Rb is a key hub integrating external mitogenic and differentiation signals to make this decision.

To appreciate this crucial role of Rb in the cell cycle, it is helpful to review the mechanisms that enforce the G_1/S transition.

- The initiation of DNA replication (S phase) requires the activity of cyclin E/CDK2 complexes, and expression of cyclin E is dependent on the E2F family of transcription factors. Early in G_1, Rb is in its hypophosphorylated active form, and it binds to and inhibits the E2F family of transcription factors, preventing transcription of cyclin E. Hypophosphorylated Rb blocks E2F-mediated transcription in at least two ways (Fig. 5–22). First, it sequesters E2F, preventing it from interacting with other transcriptional activators. Second, Rb recruits chromatin remodeling proteins, such as histone deacetylases and histone methyltransferases, which bind to the promoters of E2F-responsive genes such as cyclin E. These enzymes modify chromatin at the promoters to make DNA insensitive to transcription factors.
- This situation is changed on mitogenic signaling. Growth factor signaling leads to cyclin D expression and activation of cyclin D–CDK4/6 complexes. These complexes phosphorylate Rb, inactivating the protein and releasing E2F to induce target genes such as cyclin E. Expression of cyclin E then stimulates DNA replication and progression through the cell cycle. When the cells enter S phase, they are committed to divide without additional growth factor stimulation. During the ensuing M phase, the phosphate groups are removed from Rb by cellular phosphatases, regenerating the hypophosphorylated form of Rb.
- E2F is not the sole target of Rb. The versatile Rb protein binds to a variety of other transcription factors that regulate cell differentiation. For example, Rb stimulates myocyte-, adipocyte-, melanocyte-, and macrophage-specific transcription factors. Thus, the Rb pathway couples control of cell cycle progression at G_0-G_1 with differentiation, which may explain how differentiation is associated with exit from the cell cycle.

In view of the centrality of Rb to the control of the cell cycle, an interesting question is why *RB* is not mutated in every cancer. In fact, mutations in other genes that control Rb phosphorylation can mimic the effect of *RB* loss; such genes are mutated in many cancers that seem to have normal *RB* genes. For example, mutational activation of CDK4 or overexpression of cyclin D favors cell proliferation by facilitating Rb phosphorylation and inactivation. Indeed, cyclin D is overexpressed in many tumors because of gene amplification or translocation. Mutational inactivation of CDKIs also would drive the cell cycle by unregulated activation of cyclins and CDKs. As mentioned earlier, the *CDKN2A* gene is an extremely common target of deletion or mutational inactivation in human tumors.

The emerging paradigm is that loss of normal cell cycle control is central to malignant transformation and that at least one of the four key regulators of the cell cycle (CDKN2A, cyclin D, CDK4, Rb) is mutated in most human cancers. Furthermore, the transforming proteins of several oncogenic human DNA viruses act, in part, by neutralizing the growth inhibitory activities of Rb. For example, the human papillomavirus (HPV) E7 protein binds to the hypophosphorylated form of Rb, preventing it from inhibiting the E2F transcription factors. Thus, Rb is functionally deleted, leading to uncontrolled growth.

SUMMARY

RB Gene: Governor of the Cell Cycle

- Rb exerts antiproliferative effects by controlling the G_1-to-S transition of the cell cycle. In its active form, Rb is hypophosphorylated and binds to E2F transcription factor. This interaction prevents transcription of genes like cyclin E that are needed for DNA replication, and so the cells are arrested in G_1.
- Growth factor signaling leads to cyclin D expression, activation of the cyclin D–CDK4/6 complexes, inactivation of Rb by phosphorylation, and thus release of E2F.
- Loss of cell cycle control is fundamental to malignant transformation. Almost all cancers have a disabled G_1

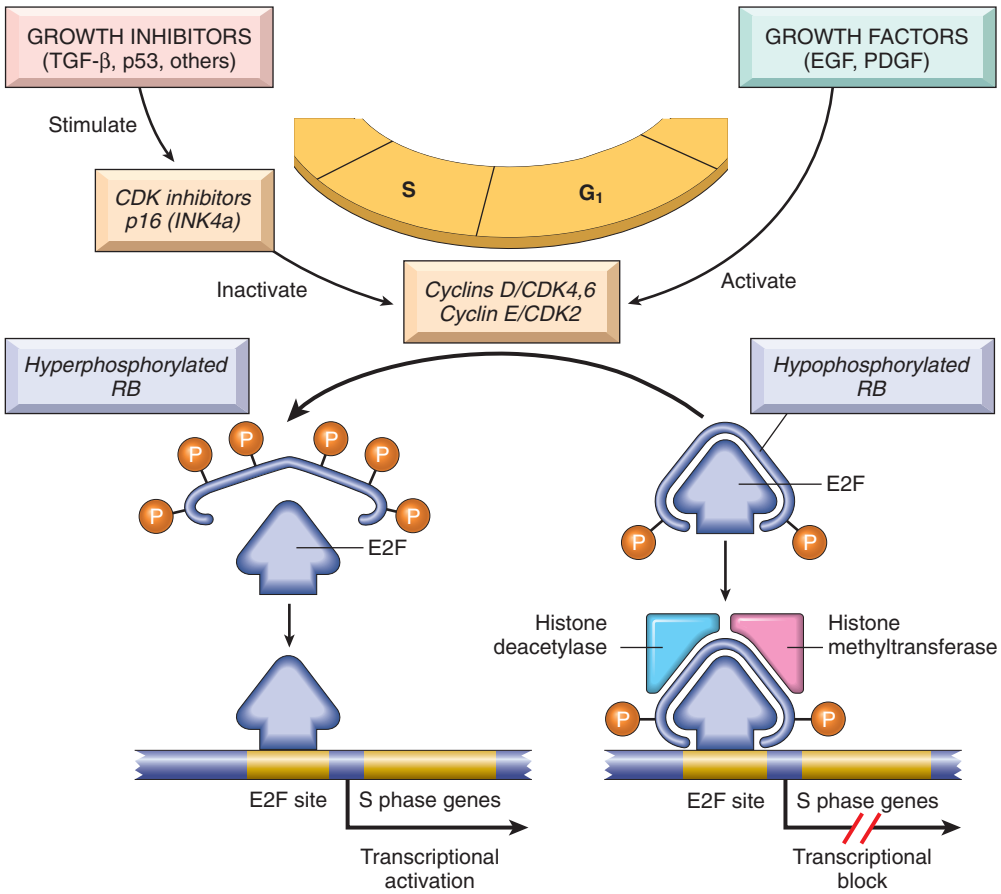


Figure 5–22 The role of Rb in regulating the G_1–S checkpoint of the cell cycle. Hypophosphorylated Rb in complex with the E2F transcription factors binds to DNA, recruits chromatin remodeling factors (histone deacetylases and histone methyltransferases), and inhibits transcription of genes whose products are required for the S phase of the cell cycle. When Rb is phosphorylated by the cyclin D–CDK4, cyclin D–CDK6, and cyclin E–CDK2 complexes, it releases E2F. The latter then activates transcription of S-phase genes. The phosphorylation of Rb is inhibited by CDKIs, because they inactivate cyclin-CDK complexes. Virtually all cancer cells show dysregulation of the G_1–S checkpoint as a result of mutation in one of four genes that regulate the phosphorylation of Rb; these genes are *RB*, *CDK4*, *cyclin D*, and *CDKN2A [p16]*. EGF, epidermal growth factor; PDGF, platelet-derived growth factor.

checkpoint due to mutation of either *RB* or genes that affect Rb function, such as cyclin D, CDK4, and CDKIs.

- Many oncogenic DNA viruses, like HPV, encode proteins (e.g., E7) that bind to Rb and render it nonfunctional.

TP53 *Gene: Guardian of the Genome*

The p53-encoding tumor suppressor gene, *TP53,* is one of the most commonly mutated genes in human cancers. The p53 protein thwarts neoplastic transformation by three interlocking mechanisms: activation of temporary cell cycle arrest (termed quiescence), induction of permanent cell cycle arrest (termed senescence), or triggering of programmed cell death (termed apoptosis). If Rb "senses" external signals, p53 can be viewed as a central monitor of internal stress, directing the stressed cells toward one of these three pathways.

A variety of stresses trigger the p53 response pathways, including anoxia, inappropriate oncoprotein activity (e.g., MYC or RAS), and damage to the integrity of DNA. By managing the DNA damage response, p53 plays a central role in maintaining the integrity of the genome, as described next.

In nonstressed, healthy cells, p53 has a short half-life (20 minutes) because of its association with MDM2, a protein that targets p53 for destruction. When the cell is stressed, for example, by an assault on its DNA, "sensors" that include protein kinases such as ATM (ataxia telangiectasia mutated) are activated. These activated complexes catalyze post-translational modifications in p53 that release it from MDM2 and increase its half-life and enhance its ability to drive the transcription of target genes. Hundreds of genes whose transcription is triggered by p53 have been found. These genes suppress neoplastic transformation by three mechanisms:

- *p53-mediated cell cycle arrest may be considered the primordial response to DNA damage* (Fig. 5–23). It occurs late in the G_1 phase and is caused mainly by p53-dependent transcription of the CDKI gene *CDKN1A (p21).* The p21 protein, as described earlier, inhibits cyclin–CDK complexes and prevents phosphorylation of Rb, thereby arresting cells in the G_1 phase. Such a pause in cell cycling is welcome, because it gives the cells "breathing time" to repair DNA damage. The p53 protein also

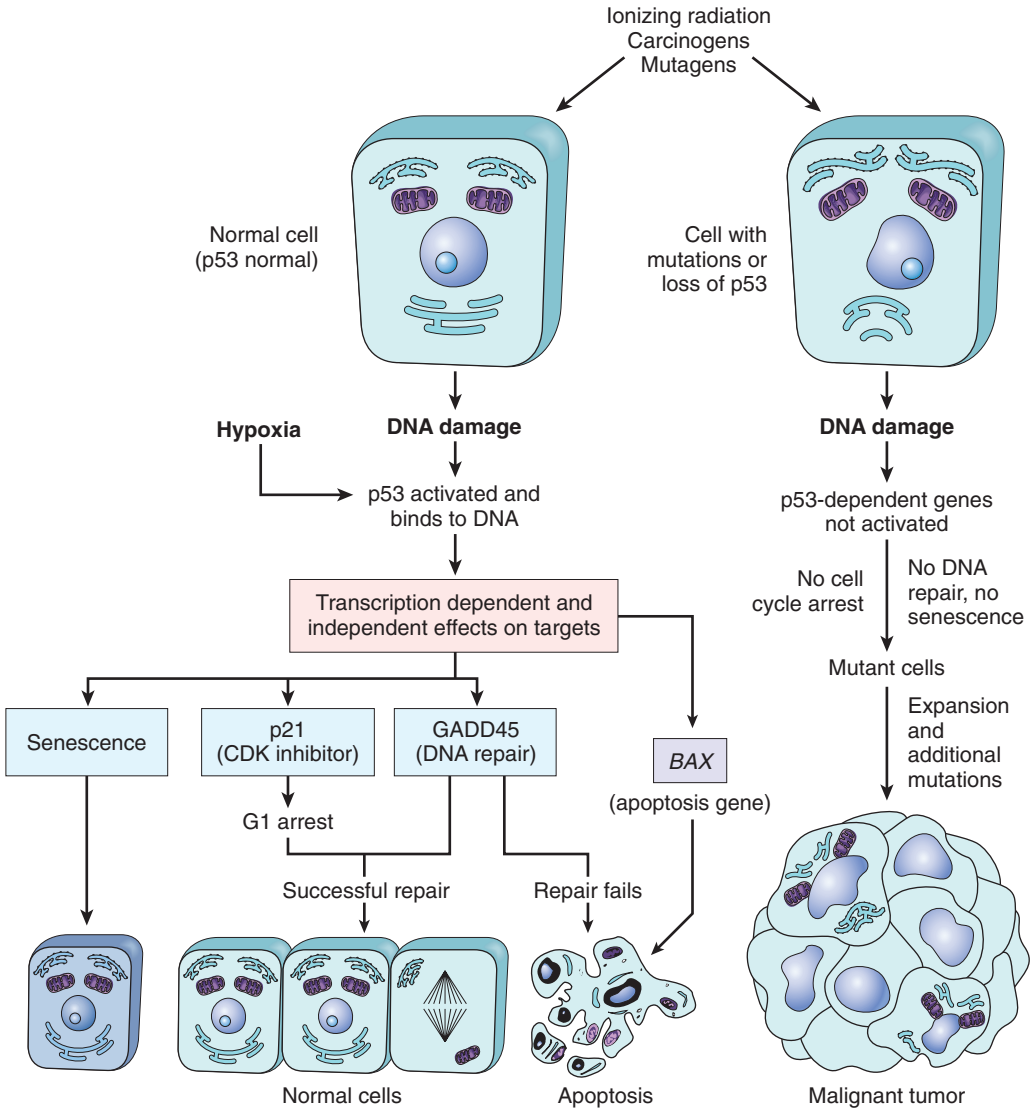


Figure 5–23 The role of p53 in maintaining the integrity of the genome. Activation of normal p53 by DNA-damaging agents or by hypoxia leads to cell cycle arrest in G_1 and induction of DNA repair, by transcriptional upregulation of the cyclin-dependent kinase inhibitor *CDKN1A* (p21) and the *GADD45* genes. Successful repair of DNA allows cells to proceed with the cell cycle; if DNA repair fails, p53 triggers either apoptosis or senescence. In cells with loss or mutations of *TP53*, DNA damage does not induce cell cycle arrest or DNA repair, and genetically damaged cells proliferate, giving rise eventually to malignant neoplasms.

induces expression of DNA damage repair genes. If DNA damage is repaired successfully, p53 upregulates transcription of MDM2, leading to destruction of p53 and relief of the cell cycle block. If the damage cannot be repaired, the cell may enter p53-induced senescence or undergo p53-directed apoptosis.

- *p53-induced senescence is a permanent cell cycle arrest* characterized by specific changes in morphology and gene expression that differentiate it from quiescence or reversible cell cycle arrest. Senescence requires activation of p53 and/or Rb and expression of their mediators, such as the CDKIs. The mechanisms of senescence are unclear but seem to involve global chromatin changes, which drastically and permanently alter gene expression.
- *p53-induced apoptosis of cells with irreversible DNA damage is the ultimate protective mechanism against neoplastic transformation.* It is mediated by several pro-apoptotic genes such as *BAX* and *PUMA* (described later).

Until recently it was thought that these functions of p53 were mediated exclusively by transcriptional activation of genes with antiproliferative, apoptotic, and senescence-inducing functions, as discussed earlier. But the waters were muddied when it was discovered that p53 represses a subset of pro-proliferative and anti-apoptotic genes as well. How could p53, a transcriptional activator, repress gene function? The answer came from the discovery that p53 can transcriptionally activate certain miRNAs (the "small guys with big clubs"). As discussed in Chapter 6, miRNAs can prevent translation of their target genes. The miRNAs activated by p53 can inhibit the translation of pro-proliferative genes such as cyclins and anti-apoptotic genes such as *BCL2*.

To summarize, p53 is activated by stresses such as DNA damage and assists in DNA repair by causing G_1 arrest and inducing DNA repair genes. A cell with damaged DNA that cannot be repaired is directed by *p53 to either enter senescence or undergo apoptosis* (Fig. 5–25). In view of these activities, *p53 has been rightfully called the "guardian of the genome."* With homozygous loss of the *TP53* gene, DNA damage goes unrepaired, mutations become fixed in dividing cells, and the cell turns onto a one-way street leading to malignant transformation.

Confirming the importance of *TP53* in controlling carcinogenesis, more than 70% of human cancers have a defect in this gene, and the remaining malignant neoplasms have defects in genes upstream or downstream of *TP53.* Biallelic loss of the *TP53* gene is found in virtually every type of cancer, including carcinomas of the lung, colon, and breast—the three leading causes of cancer deaths. In most cases, inactivating mutations affecting both *TP53* alleles are acquired in somatic cells. Less commonly, some patients inherit a mutant *TP53* allele; the resulting disease is called the *Li-Fraumeni syndrome.* As with the *RB* gene, inheritance of one mutant allele predisposes affected persons to develop malignant tumors because only one additional hit is needed to inactivate the second, normal allele. Patients with the Li-Fraumeni syndrome have a 25-fold greater chance of developing a malignant tumor by age 50 compared with the general population. In contrast with tumors developing in patients who inherit a mutant *RB* allele, the spectrum of tumors that develop in patients with the Li-Fraumeni syndrome is varied; the most common types are sarcomas, breast cancer, leukemia, brain tumors, and carcinomas of the adrenal cortex. Compared with persons diagnosed with sporadic tumors, patients with Li-Fraumeni syndrome develop tumors at a younger age and may develop multiple primary tumors.

As with Rb protein, normal p53 also can be rendered nonfunctional by certain DNA viruses. Proteins encoded by oncogenic HPVs, hepatitis B virus (HBV), and possibly Epstein-Barr virus (EBV) can bind to normal p53 and nullify its protective function. Thus, DNA viruses can subvert two of the best-understood tumor suppressors, Rb and p53.

SUMMARY

TP53 Gene: Guardian of the Genome

- The p53 protein is the central monitor of stress in the cell and can be activated by anoxia, inappropriate oncogene signaling, or DNA damage. Activated p53 controls the expression and activity of genes involved in cell cycle arrest, DNA repair, cellular senescence, and apoptosis.
- DNA damage leads to activation of p53 by phosphorylation. Activated p53 drives transcription of *CDKN1A (p21)*, which prevents Rb phosphorylation, thereby causing a G_1-S block in the cell cycle. This pause allows the cells to repair DNA damage.
- If DNA damage cannot be repaired, p53 induces cellular senescence or apoptosis.
- Of human tumors, 70% demonstrate biallelic loss of *TP53*. Patients with the rare Li-Fraumeni syndrome inherit one defective copy in the germ line and lose the second one in somatic tissues; such persons develop a variety of tumors.
- As with Rb, p53 can be incapacitated by binding to proteins encoded by oncogenic DNA viruses such as HPV.

Transforming Growth Factor-β Pathway

Although much is known about the circuitry that applies brakes to the cell cycle, the molecules that transmit antiproliferative signals to cells are less well characterized. Best-known is TGF-β, a member of a family of dimeric growth factors that includes bone morphogenetic proteins and activins. In most normal epithelial, endothelial, and hematopoietic cells, TGF-β is a potent inhibitor of proliferation. It regulates cellular processes by binding to a complex composed of TGF-β receptors I and II. Dimerization of the receptor upon ligand binding leads to a cascade of events that result in the transcriptional activation of CDKIs with growth-suppressing activity, as well as repression of growth-promoting genes such as *MYC, CDK2, CDK4,* and those encoding cyclins A and E.

In many forms of cancer, the growth-inhibiting effects of the TGF-β pathways are impaired by mutations affecting TGF-β signaling. These mutations may alter the type II TGF-β receptor or SMAD molecules that serve to transduce antiproliferative signals from the receptor to the nucleus. Mutations affecting the type II receptor are seen in cancers of the colon, stomach, and endometrium. Mutational inactivation of SMAD4, 1 of the 10 proteins known to be involved in TGF-β signaling, is common in pancreatic cancers. *In 100% of pancreatic cancers and 83% of colon cancers, at least one component of the TGF-β pathway is mutated.* In many cancers, however, loss of TGF-β–mediated growth control occurs at a level downstream of the core signaling pathway, for example, loss of p21 and/or persistent expression of MYC. These tumor cells can then use other elements of the TGF-β–induced program, including immune system suppression–evasion or promotion of angiogenesis, to facilitate tumor progression. Thus, TGF-β can function to prevent or promote tumor growth, depending on the state of other genes in the cell. Indeed, in many late-stage tumors, TGF-β signaling activates epithelial-to-mesenchymal transition (EMT), a process that promotes migration, invasion, and metastasis, as described later.

Contact Inhibition, NF2, *and APC*

When nontransformed cells are grown in culture, they proliferate until confluent monolayers are generated; cell–cell contacts formed in these monolayers suppress further cell proliferation. Of importance, "contact inhibition" is abolished in cancer cells, allowing them to pile on top of one another. The mechanisms that govern contact inhibition are only now being discovered. Cell–cell contacts in many tissues are mediated by homodimeric interactions between transmembrane proteins called cadherins. E-cadherin (E for epithelial) mediates cell–cell contact in epithelial layers. How E-cadherin maintains normal contact inhibition is not fully understood. One mechanism that sustains contact inhibition is mediated by the tumor suppressor gene *NF2*. Its product, neurofibromin-2, more commonly called

merlin, facilitates E-cadherin mediated contact inhibition. Homozygous loss of *NF2* is known to cause a form of neural tumors associated with the condition called neurofibromatosis.

There are other mechanisms of E-cadherin regulation as well. One such mechanism is illustrated by the rare hereditary disease *adenomatous polyposis coli* (APC). This disorder is characterized by the development of numerous adenomatous polyps in the colon that have a very high incidence of transformation into colonic cancers. They consistently show loss of a tumor suppressor gene called *APC* (named for the disease). The *APC* gene exerts antiproliferative effects in an unusual manner. It encodes a cytoplasmic protein whose dominant function is to regulate the intracellular levels of β-catenin, a protein with many functions. On the one hand, β-catenin binds to the cytoplasmic portion of E-cadherin; on the other hand, it can translocate to the nucleus and activate cell proliferation. Here the focus is on the latter function of this protein. β-Catenin is an important component of the so-called WNT signaling pathway that regulates cell proliferation (illustrated in Fig. 5–24). WNT is a soluble factor that can induce cellular proliferation. It does so by binding to its receptor and transmitting signals that prevent the degradation of β-catenin, allowing it to translocate to the nucleus, where it acts as a transcriptional activator in conjunction with another molecule, called TcF (Fig. 5–24, *B*). In quiescent cells, which are not exposed to WNT, cytoplasmic β-catenin is degraded by a *destruction complex,* of which APC is an integral part (Fig. 5–24, *A*). With loss of APC (in malignant cells), β-catenin degradation is prevented, and the WNT signaling response is inappropriately activated in the absence of WNT (Fig. 5–24, *C*). This leads to transcription of growth-promoting genes, such as cyclin D1 and *MYC,* as well as transcriptional regulators, such as TWIST and SLUG, that repress E-cadherin expression and thus reduce contact inhibition.

APC behaves as a typical tumor suppressor gene. Persons born with one mutant allele typically are found to have hundreds to thousands of adenomatous polyps in the colon by their teens or 20s; these polyps show loss of the other *APC* allele. Almost invariably, one or more polyps undergo malignant transformation, as discussed later. *APC* mutations are seen in 70% to 80% of sporadic colon cancers. Colonic cancers that have normal *APC* genes show activating mutations of β-catenin that render them refractory to the degrading action of APC.

SUMMARY

Transforming Growth Factor-β and APC–β-Catenin Pathways

- TGF-β inhibits proliferation of many cell types by activation of growth-inhibiting genes such as CDKIs and suppression of growth-promoting genes such as *MYC* and those encoding cyclins.

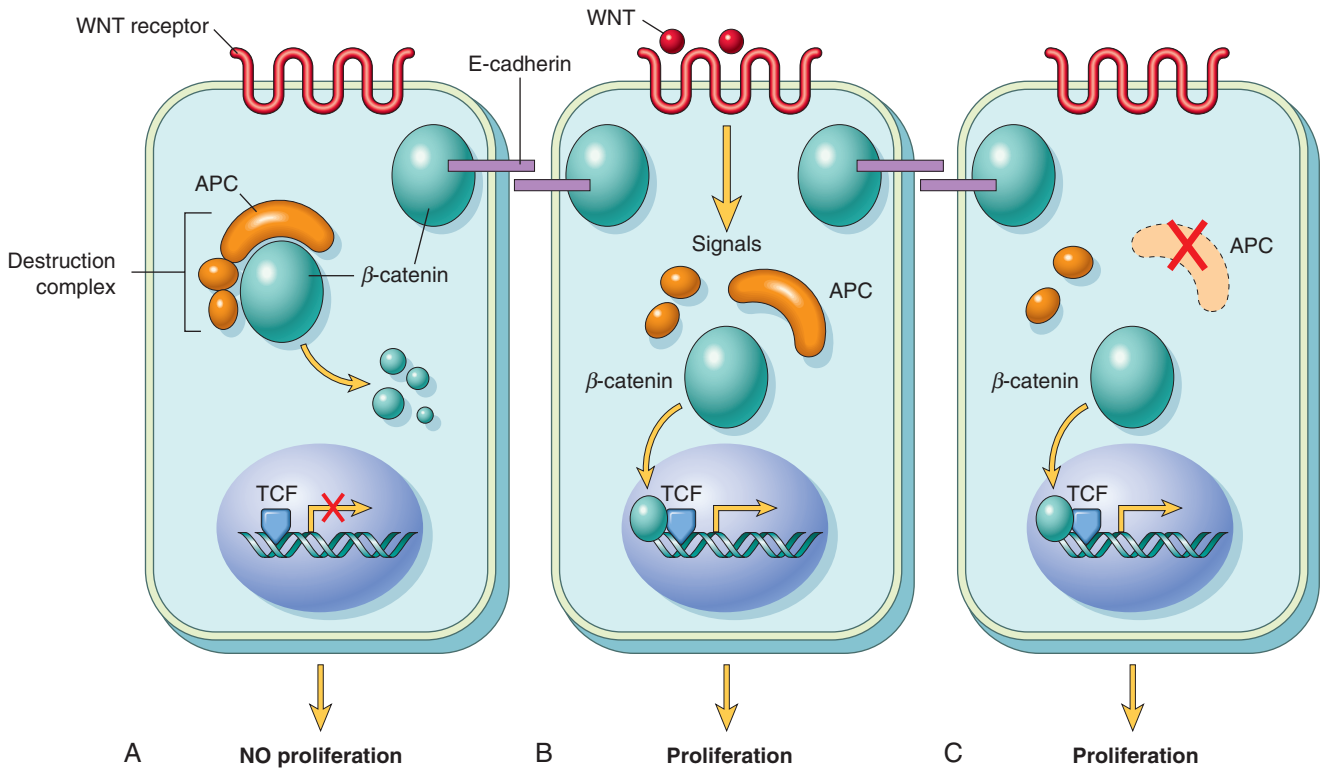


Figure 5–24 A–C, The role of APC in regulating the stability and function of β-catenin. APC and β-catenin are components of the WNT signaling pathway. In resting cells (not exposed to WNT), β-catenin forms a macromolecular complex containing the APC protein. This complex leads to the destruction of β-catenin, and intracellular levels of β-catenin are low. When cells are stimulated by secreted WNT molecules, the *destruction complex* is deactivated, β-catenin degradation does not occur, and cytoplasmic levels increase. β-Catenin translocates to the nucleus, where it binds to TCF, a transcription factor that activates several genes involved in the cell cycle. When APC is mutated or absent, the destruction of β-catenin cannot occur. β-Catenin translocates to the nucleus and coactivates genes that promote the cell cycle, and cells behave as if they are under constant stimulation by the WNT pathway.

- TGF-β function is compromised in many tumors by mutations in its receptors (colon, stomach, endometrium) or by mutational inactivation of *SMAD* genes that transduce TGF-β signaling (pancreas).
- E-cadherin maintains contact inhibition, which is lost in malignant cells.
- *APC* gene exerts antiproliferative actions by regulating the destruction of the cytoplasmic protein β-catenin. With a loss of *APC*, β-catenin is not destroyed, and it translocates to the nucleus, where it acts as a growth-promoting transcription factor.
- In familial adenomatous polyposis syndrome, inheritance of a germ line mutation in the *APC* gene and sporadic loss of the sole normal allele causes the development of hundreds of colonic polyps at a young age. Inevitably, one or more of these polyps evolves into a colonic cancer. Somatic loss of both alleles of the *APC* gene is seen in approximately 70% of sporadic colon cancers.

Evasion of Cell Death

As discussed in Chapter 1, apoptosis, or programmed cell death, refers to an orderly dismantling of cells into component pieces that can then be consumed and disposed of by neighboring cells. *It is now well established that accumulation of neoplastic cells may result not only from activation of growth-promoting oncogenes or inactivation of growth-suppressing tumor suppressor genes but also from mutations in the genes that regulate apoptosis.*

The apoptotic pathway can be divided into upstream regulators and downstream effectors. The regulators are divided into two major pathways, one interpreting extracellular or extrinsic signals and the other interpreting intracellular signals. Stimulation of either pathway results in activation of a normally inactive protease (caspase-8 or caspase-9, respectively), which initiates a proteolytic cascade involving "executioner" caspases that disassemble the cell in orderly fashion. The cellular remains are then efficiently consumed by the cellular neighbors and professional phagocytes, without stimulating inflammation. Figure 5–25 shows, in simplified form, the sequence of events that lead to apoptosis by signaling through death receptors, which are members of the TNF receptor family (extrinsic pathway), and by DNA damage and other stresses (intrinsic pathway).

- The extrinsic (death receptor) pathway is initiated when a TNF receptor, such as CD95 (Fas), is bound to its ligand, CD95L, leading to trimerization of the receptor and its cytoplasmic *death domains,* which attract the intracellular adaptor protein FADD. This protein recruits procaspase-8 to form the death-inducing signaling complex. Procaspase-8 is activated by cleavage into smaller subunits, generating caspase-8. Caspase-8 then activates downstream caspases such as caspase-3, an *executioner caspase* that cleaves DNA and other substrates to cause cell death.
- The intrinsic (mitochondrial) pathway of apoptosis is triggered by a variety of stimuli, including withdrawal of survival factors, stress, and injury. Activation of this pathway leads to permeabilization of the mitochondrial outer membrane and release of molecules, such as cytochrome c, that initiate apoptosis.

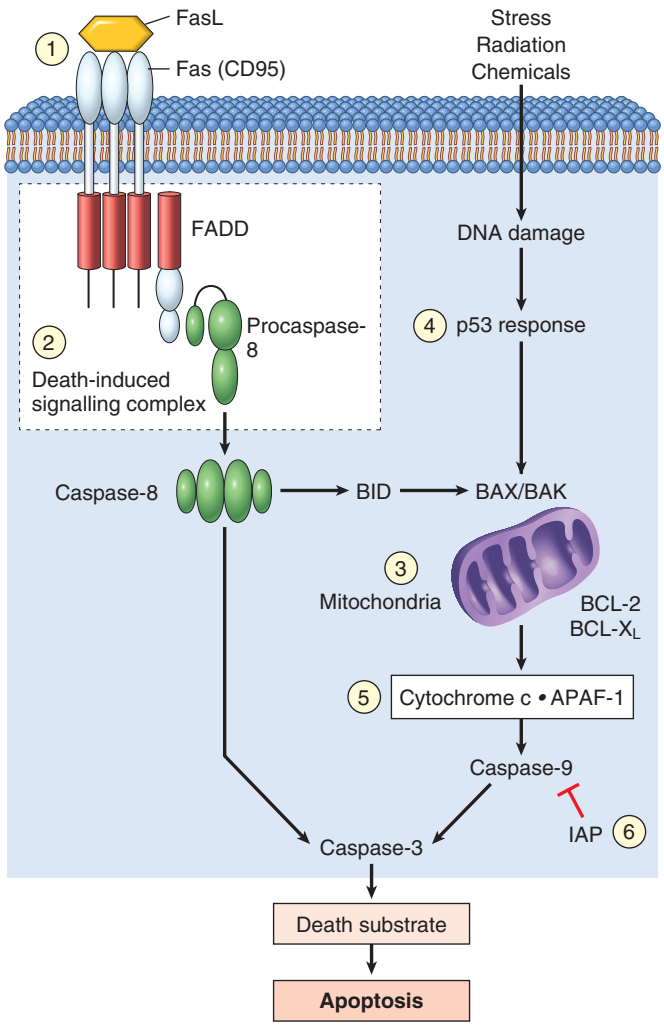


Figure 5–25 Simplified schema of CD95 receptor–induced and DNA damage–triggered pathways of apoptosis and mechanisms used by tumor cells to evade cell death: *1*, Reduced CD95 level. *2*, Inactivation of death-induced signaling complex by FLICE protein. *3*, Reduced egress of cytochrome c from mitochondrion as a result of upregulation of BCL2. *4*, Reduced levels of pro-apoptotic BAX resulting from loss of p53. *5*, Loss of APAF-1. *6*, Upregulation of inhibitors of apoptosis.

The integrity of the mitochondrial outer membrane is regulated by pro-apoptotic and anti-apoptotic members of the BCL2 family of proteins. The pro-apoptotic proteins BAX and BAK are required for apoptosis and directly promote mitochondrial permeabilization. Their action is inhibited by the anti-apoptotic members of this family exemplified by BCL2 and BCL-X_L. A third set of proteins, the so-called BH3-only proteins, which include BAD, BID, and PUMA, regulate the balance between the pro- and anti-apoptotic members of the BCL2 family. The BH3-only proteins promote apoptosis by neutralizing the actions of anti-apoptotic proteins like BCL2 and BCL-X_L. When the sum total of all BH3 proteins expressed "overwhelms" the anti-apoptotic BCL2/BCLX_L protein barrier, BAX and BAK are activated and form pores in the mitochondrial membrane. Cytochrome c leaks into the cytosol, where it binds to APAF-1 and

activates caspase-9. Like caspase-8 of the extrinsic pathway, caspase-9 can cleave and activate the executioner caspases. Caspases can be inhibited by a family of proteins called inhibitor of apoptosis proteins (IAPs). Because of the pro-apoptotic effect of BH3 only proteins, efforts are underway to develop BH3 mimetic drugs to promote death of tumor cells.

Within this framework, it is possible to illustrate the multiple sites at which apoptosis is frustrated by cancer cells (Fig. 5–25). Of these candidates, perhaps *best-established is the role of BCL2 in protecting tumor cells from apoptosis.* Approximately 85% of B cell lymphomas of the follicular type (Chapter 11) carry a characteristic t(14;18) (q32;q21) translocation. As noted earlier, 14q32, the chromosomal locus for immunoglobulin heavy-chain genes, also is involved in the pathogenesis of Burkitt lymphoma. Juxtaposition of this transcriptionally active locus with *BCL2* (located at 18q21) causes overexpression of the BCL2 protein. This overabundance in turn increases the BCL2/BCL-X_L buffer, protecting lymphocytes from apoptosis and allowing them to survive for long periods; there is therefore a steady accumulation of B lymphocytes, resulting in lymphadenopathy and marrow infiltration. Because BCL2-overexpressing lymphomas arise in large part through reduced cell death rather than explosive cell proliferation, they tend to be indolent (slow-growing) compared to other lymphomas. In some instances, reduced levels of CD95 may render the tumor cells less susceptible to apoptosis by Fas ligand (FasL). Some tumors have high levels of FLIP, a protein that can bind death-inducing signaling complex and prevent activation of caspase 8.

As mentioned previously, *TP53 is an important pro-apoptotic gene that induces apoptosis in cells that are unable to repair DNA damage.* Similarly, unrestrained action of growth-promoting genes such as *MYC* also leads to apoptosis. Thus, both major oncogenic pathways—inability to repair DNA damage and inappropriate activation of oncogenes—converge on the apoptotic machinery, which, by causing cell death, acts as a major barrier to carcinogenesis.

Autophagy

As described in Chapter 1, autophagy is a key catabolic process that helps balance synthesis, degradation, and recycling of cellular products. During autophagy, cellular organelles, such as ribosomes and mitochondria, are sequestered from the rest of the cell by a membrane (autophagosome) and then fused to a lysosome, where they are degraded and utilized for cellular energy generation. The same process can signal cells to die if they cannot be rescued by the recycling of organelles. It is a tightly regulated process that plays an important role in normal cell function, and can help starving cells shift nutrients from unused cell processes to vital ones. Autophagy, like apoptosis, has regulatory and effector machinery. The effector components consist of proteins that lead to the formation of autophagosomes and direct their contents to lysosomes. Not surprisingly, the regulatory components of autophagy overlap with many of the signaling components that regulate apoptosis. For example, a protein, Beclin-1, required for autophagy, belongs to the BH3 domain containing proteins that regulate apoptosis. When cells sense internal stress (e.g., DNA damage), they may undergo apoptosis or Beclin-1–induced autophagy. Thus, autophagy, by analogy with apoptosis, appears to prevent the growth of tumor cells. Later in tumor growth, however, autophagy may be helpful to tumors. The metabolites generated by autophagy may supply crucial building blocks for growth and survival in the nutrient-poor environments that tumor cells inhabit. Indeed, autophagy may promote tumor survival in unfriendly climates or during therapy. Thus, autophagy may act as either a "friend" or a "foe," depending on other internal and external factors.

SUMMARY

Evasion of Apoptosis

- Apoptosis can be initiated through extrinsic or intrinsic pathways.
- Both pathways result in the activation of a proteolytic cascade of caspases that destroys the cell.
- Mitochondrial outer membrane permeabilization is regulated by the balance between pro-apoptotic (e.g., BAX, BAK) and anti-apoptotic molecules (BCL2, BCL-X_L). BH-3–only molecules activate apoptosis by tilting the balance in favor of the pro-apoptotic molecules.
- In 85% of follicular B cell lymphomas, the anti-apoptotic gene *BCL2* is activated by the t(14;18) translocation.
- Stress may also induce cells to consume their components in a process called autophagy. Cancer cells may accumulate mutations to avoid autophagy, or may corrupt the process to provide parts for continued growth.

Limitless Replicative Potential

As discussed previously in the context of cellular aging (Chapter 1), most normal human cells have a capacity of 60 to 70 doublings. Thereafter, the cells lose the capacity to divide and enter senescence. This phenomenon has been ascribed to progressive shortening of *telomeres* at the ends of chromosomes. The consequences of such shortening, when pronounced, are drastic:

- Short telomeres seem to be recognized by the DNA repair machinery as double-stranded DNA breaks, leading to cell cycle arrest and senescence, mediated by *TP53* and *RB*. In cells in which the checkpoints are disabled by *TP53* or *RB* mutations, the nonhomologous end-joining pathway is activated in a last-ditch effort to save the cell, joining the shortened ends of two chromosomes.
- Such an inappropriately activated repair system results in dicentric chromosomes that are pulled apart at anaphase, resulting in new double-stranded DNA breaks. The resulting genomic instability from the repeated bridge–fusion–breakage cycles eventually produces mitotic catastrophe, characterized by massive apoptosis.

It follows that for tumors to grow indefinitely, as they often do, loss of growth restraints is not enough. Tumor cells also must develop ways to avoid both cellular senescence and mitotic

catastrophe (Fig. 5–26). If during crisis a cell manages to reactivate telomerase, the bridge–fusion–breakage cycles cease, and the cell is able to avoid death. However, during this period of genomic instability that precedes telomerase activation, numerous mutations could accumulate, helping the cell march toward malignancy. Telomerase, active in normal stem cells, normally is absent from, or present at very low levels in, most somatic cells. By contrast, telomere maintenance is seen in virtually all types of cancers. In 85% to 95% of cancers, this is due to upregulation of the enzyme telomerase. A few tumors use other mechanisms, termed alternative lengthening of telomeres, which probably depend on DNA recombination.

Of interest, in a study of the progression from colonic adenoma to colonic adenocarcinoma, early lesions had a high degree of genomic instability with low telomerase expression, whereas malignant lesions had complex karyotypes with high levels of telomerase activity, consistent with a model of telomere-driven tumorigenesis in human cancer. Thus, it appears that in this model, unregulated proliferation in incipient tumors leads to telomere shortening, followed by chromosomal instability and mutation accumulation. If telomerase is then reactivated in these cells, telomeres are extended and these mutations become fixed, contributing to tumor growth. Several other mechanisms of genomic instability are discussed later.

SUMMARY

Limitless Replicative Potential

- In normal cells, which lack expression of telomerase, the shortened telomeres generated by cell division eventually activate cell cycle checkpoints, leading to senescence and placing a limit on the number of divisions a cell may undergo.
- In cells that have disabled checkpoints, DNA repair pathways are inappropriately activated by shortened telomeres, leading to massive chromosomal instability and mitotic crisis.
- Tumor cells reactivate telomerase, thus staving off mitotic catastrophe and achieving immortality.

Development of Sustained Angiogenesis

Even with all the growth advantages, as described previously, tumors cannot enlarge beyond 1 to 2 mm in diameter unless they are vascularized. Like normal tissues, tumors require delivery of oxygen and nutrients and removal of waste products; the 1- to 2-mm zone presumably represents the maximal distance across which oxygen, nutrients, and waste can diffuse from blood vessels. Cancer

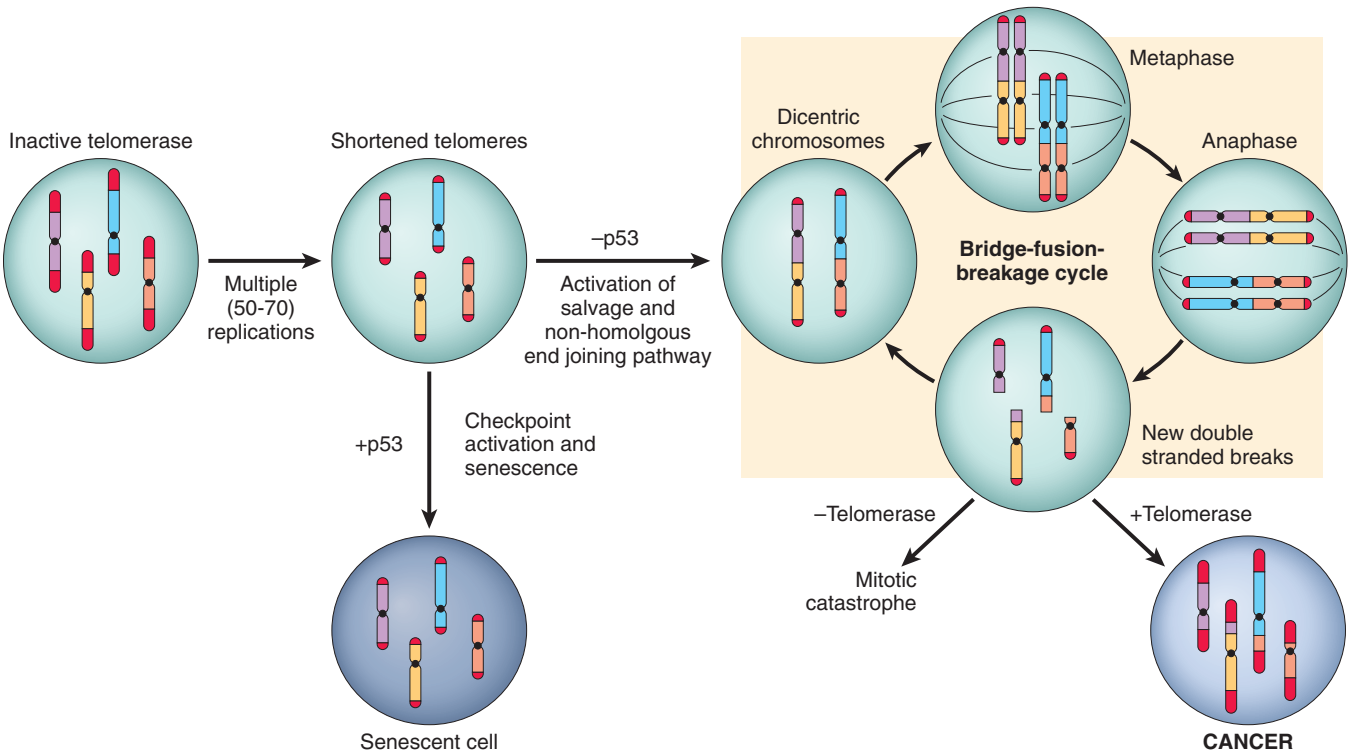


Figure 5–26 Sequence of events in the development of limitless replicative potential. Replication of somatic cells, which do not express telomerase, leads to shortened telomeres. In the presence of competent checkpoints, cells undergo arrest and enter nonreplicative senescence. In the absence of checkpoints, DNA repair pathways are inappropriately activated, leading to the formation of dicentric chromosomes. At mitosis, the dicentric chromosomes are pulled apart, generating random double-stranded breaks, which then activate DNA repair pathways, leading to the random association of double-stranded ends and the formation, again, of dicentric chromosomes. Cells undergo numerous rounds of this bridge–fusion–breakage cycle, which generates massive chromosomal instability and numerous mutations. If cells fail to reexpress telomerase, they eventually undergo mitotic catastrophe and death. Reexpression of telomerase allows the cells to escape the bridge–fusion–breakage cycle, thus promoting their survival and tumorigenesis.

cells (and large benign tumors) can stimulate neoangiogenesis, during which new vessels sprout from previously existing capillaries, or, in some cases, vasculogenesis, in which endothelial cells are recruited from the bone marrow. Tumor vasculature is abnormal, however. The vessels are leaky and dilated, with a haphazard pattern of connection. Neovascularization has a dual effect on tumor growth: Perfusion supplies needed nutrients and oxygen, and newly formed endothelial cells stimulate the growth of adjacent tumor cells by secreting growth factors, such as insulin-like growth factors, PDGF, and granulocyte-macrophage colony-stimulating factor. Angiogenesis is required not only for continued tumor growth but also for access to the vasculature and hence for metastasis. *Angiogenesis is thus a necessary biologic correlate of neoplasia, both benign and malignant.*

How do growing tumors develop a blood supply? The emerging paradigm is that tumor angiogenesis is controlled by a balance between pro-angiogenic and inhibitory factors.

- The prototypical angiogenesis inducer and inhibitor are vascular endothelial growth factor (VEGF) and thrombospondin-1 (TSP-1), respectively. Early in their growth, most human tumors do not induce angiogenesis. They remain small or in situ for years until the angiogenic switch terminates this stage of vascular quiescence. Normal p53 induces synthesis of TSP-1.
- The molecular basis of the angiogenic switch involves increased production of angiogenic factors and/or loss of angiogenesis inhibitors. These factors may be produced directly by the tumor cells themselves or by inflammatory cells (e.g., macrophages) or other stromal cells associated with the tumors.
- Proteases, elaborated either by the tumor cells directly or from stromal cells in response to the tumor, also are involved in regulating the balance between angiogenic and anti-angiogenic factors. Many proteases can release the angiogenic basic FGF stored in the extracellular matrix (ECM); conversely, three potent angiogenesis inhibitors—angiostatin, endostatin, and vasculostatin—are produced by proteolytic cleavage of plasminogen, collagen, and transthyretin, respectively. TSP-1, on the other hand, is produced by stromal fibroblasts themselves in response to signals from the tumor cells.
- The angiogenic switch is controlled by several physiologic stimuli, such as hypoxia. Relative lack of oxygen stimulates production of a variety of pro-angiogenic cytokines, such as vascular endothelial growth factor (VEGF), through activation of hypoxia-inducible factor-1α (HIF-1α), an oxygen-sensitive transcription factor. HIF-1α is continuously produced, but in normoxic settings the von Hippel–Lindau protein (VHL) binds to HIF-1α, leading to ubiquitination and destruction of HIF-1α.
- In hypoxic conditions, such as in a tumor that has reached a critical size, the lack of oxygen prevents HIF-1α recognition by VHL, and it is not destroyed. HIF-1α translocates to the nucleus and activates transcription of its target genes, such as VEGF. Because of these activities, *VHL* acts as a tumor suppressor gene, and germline mutations of the *VHL* gene are associated with hereditary renal cell cancers, pheochromocytomas, hemangiomas of the central nervous system, retinal angiomas, and renal cysts (*VHL syndrome*).
- VEGF also increases the expression of ligands that activate the Notch signaling pathway, which regulates the branching and density of the new vessels. Because of the crucial role of angiogenesis in tumor growth, much interest is focused on anti-angiogenesis therapy. Indeed, anti-VEGF antibody is now approved for the treatment of several types of cancers.

SUMMARY

Development of Sustained Angiogenesis

- Vascularization of tumors is essential for their growth and is controlled by the balance between angiogenic and anti-angiogenic factors that are produced by tumor and stromal cells.
- Hypoxia triggers angiogenesis through the actions of HIF-1α on the transcription of the pro-angiogenic factor VEGF. Because of its ability to degrade HIF-1α and thereby prevent angiogenesis, VHL acts as a tumor suppressor. Inheritance of germ line mutations of *VHL* causes VHL syndrome, characterized by the development of a variety of tumors.
- Many other factors regulate angiogenesis; for example, p53 induces synthesis of the angiogenesis inhibitor TSP-1.

Ability to Invade and Metastasize

The spread of tumors is a complex process involving a series of sequential steps called the invasion–metastasis cascade (Fig. 5–27). These steps consist of local invasion, intravasation into blood and lymph vessels, transit through the vasculature, extravasation from the vessels, formation of micrometastases, and growth of micrometastases into macroscopic tumors. Predictably, this sequence of steps may be interrupted at any stage by either host-related or tumor-related factors. For the purpose of discussion, the metastatic cascade can be subdivided into two phases: (1) invasion of ECM and (2) vascular dissemination and homing of tumor cells.

Invasion of Extracellular Matrix (ECM)

As is well recognized, human tissues are organized into a series of compartments separated from each other by two types of ECM: basement membranes and interstitial connective tissue (Chapter 2). Although organized differently, each type of ECM is composed of collagens, glycoproteins, and proteoglycans. Tumor cells must interact with the ECM at several stages in the metastatic cascade (Fig. 5–27). A carcinoma first must breach the underlying basement membrane, then traverse the interstitial connective tissue, and ultimately gain access to the circulation by penetrating the vascular basement membrane. This cycle is repeated when tumor cell emboli extravasate at a distant site. Thus, to metastasize, a tumor cell must cross several different basement membranes, as well as negotiate its way through at least two interstitial matrices. Invasion of the ECM is an active process that requires four steps (Fig. 5–28):

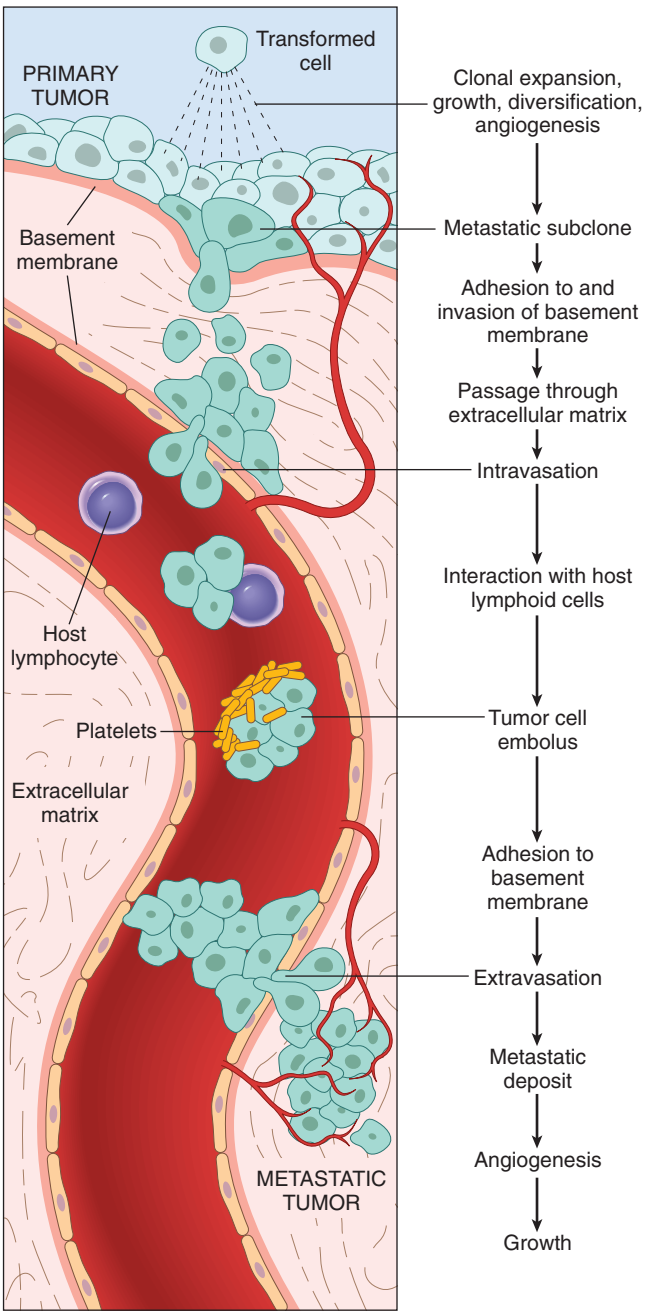


Figure 5–27 The metastatic cascade: The sequential steps involved in the hematogenous spread of a tumor.

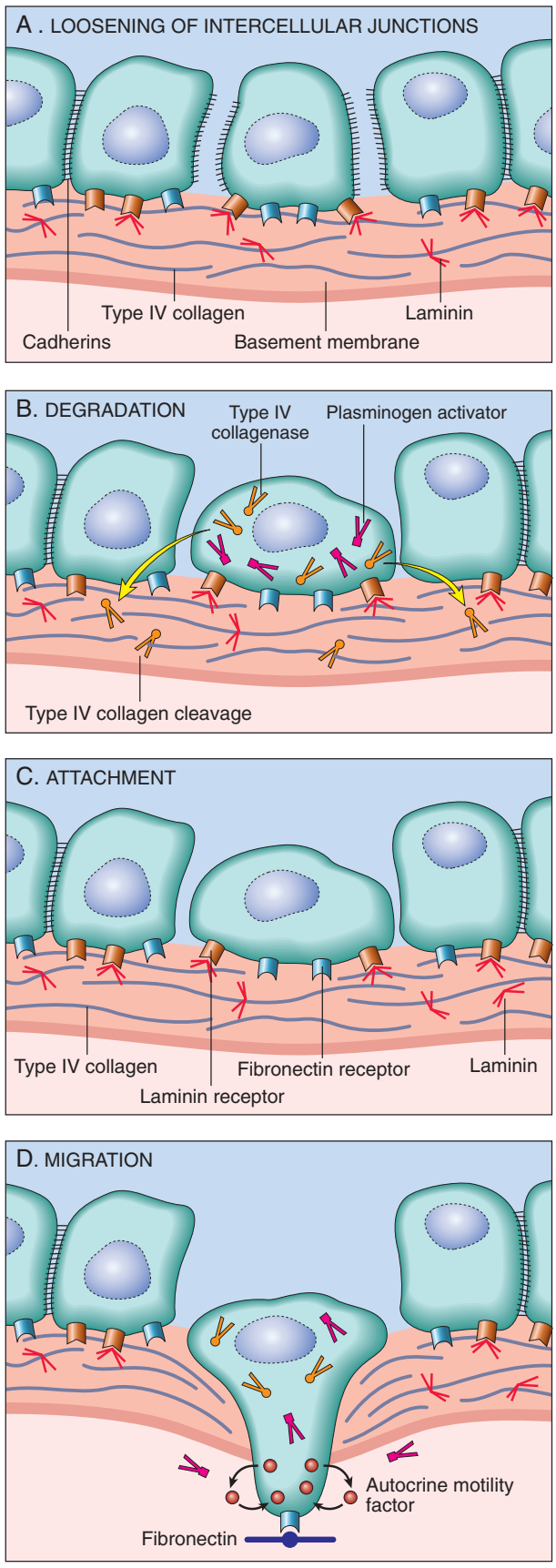


Figure 5–28 A–D, Sequence of events in the invasion of epithelial basement membranes by tumor cells. Tumor cells detach from each other because of reduced adhesiveness, then secrete proteolytic enzymes, degrading the basement membrane. Binding to proteolytically generated binding sites and tumor cell migration follow.

- The first step in the metastatic cascade is a *loosening* of tumor cells. As mentioned earlier, E-cadherins act as intercellular glues, and their cytoplasmic portions bind to β-catenin (Fig. 5–24). Adjacent E-cadherin molecules keep the cells together; in addition, as discussed earlier, E-cadherin can transmit antigrowth signals by sequestering β-catenin. *E-cadherin function is lost in almost all epithelial cancers, either by mutational inactivation of E-cadherin genes, by activation of β-catenin genes, or by inappropriate expression of the SNAIL and TWIST transcription factors, which suppress E-cadherin expression.*
- The second step in invasion is local *degradation of the basement membrane and interstitial connective tissue*. Tumor cells may either secrete proteolytic enzymes themselves or induce stromal cells (e.g., fibroblasts and inflammatory cells) to elaborate proteases. Multiple different families of proteases, such as matrix metalloproteinases (MMPs), cathepsin D, and urokinase plasminogen activator, have been implicated in tumor cell invasion. MMPs regulate tumor invasion not only by remodeling insoluble components of the basement membrane and interstitial matrix but also by releasing ECM-sequestered growth factors. Indeed, cleavage products of collagen and proteoglycans also have chemotactic, angiogenic, and growth-promoting effects. For example, MMP-9 is a gelatinase that cleaves type IV collagen of the epithelial and vascular basement membrane and also stimulates release of VEGF from ECM-sequestered pools. Benign tumors of the breast, colon, and stomach show little type IV collagenase activity, whereas their malignant counterparts overexpress this enzyme. Concurrently, the levels of metalloproteinase inhibitors are reduced so that the balance is tilted greatly toward tissue degradation. Indeed, overexpression of MMPs and other proteases has been reported for many tumors.
- The third step in invasion involves *changes in attachment of tumor cells to ECM proteins.* Normal epithelial cells have receptors, such as integrins, for basement membrane laminin and collagens that are polarized at their basal surface; these receptors help to maintain the cells in a resting, differentiated state. Loss of adhesion in normal cells leads to induction of apoptosis, while, not surprisingly, tumor cells are resistant to this form of cell death. Additionally, the matrix itself is modified in ways that promote invasion and metastasis. For example, cleavage of the basement membrane proteins, collagen IV and laminin, by MMP-2 or MMP-9 generates novel sites that bind to receptors on tumor cells and stimulate migration.
- *Locomotion* is the final step of invasion, propelling tumor cells through the degraded basement membranes and zones of matrix proteolysis. Migration is a complex, multistep process that involves many families of receptors and signaling proteins that eventually impinge on the actin cytoskeleton. Such movement seems to be potentiated and directed by tumor cell–derived cytokines, such as autocrine motility factors. In addition, cleavage products of matrix components (e.g., collagen, laminin) and some growth factors (e.g., insulin-like growth factors I and II) have chemotactic activity for tumor cells. Stromal cells also produce paracrine effectors of cell motility, such as hepatocyte growth factor/scatter factor (HGF/SCF), which binds to receptors on tumor cells. Concentrations of HGF/SCF are elevated at the advancing edges of the highly invasive brain tumor glioblastoma multiforme, supporting their role in motility.

More recently, it has become clear that the stromal cells surrounding tumor cells do not merely present a static barrier for tumor cells to traverse but rather constitute a variable environment in which reciprocal signaling between tumor cells and stromal cells may promote or prevent tumorigenesis. Stromal cells that interact with tumors include innate and adaptive immune cells (discussed later), as well as fibroblasts. A variety of studies have demonstrated that tumor-associated fibroblasts exhibit altered expression of genes that encode ECM molecules, proteases, protease inhibitors, and various growth factors. Thus, tumor cells live in a complex and ever-changing milieu composed of ECM, growth factors, fibroblasts, and immune cells, with significant cross-talk among all the components. The most successful tumors may be those that can co-opt and adapt this environment to their own nefarious ends.

Vascular Dissemination and Homing of Tumor Cells

When in the circulation, tumor cells are vulnerable to destruction by host immune cells (discussed later). In the bloodstream, some tumor cells form emboli by aggregating and adhering to circulating leukocytes, particularly platelets; aggregated tumor cells are thus afforded some protection from the antitumor host effector cells. Most tumor cells, however, circulate as single cells. Extravasation of free tumor cells or tumor emboli involves adhesion to the vascular endothelium, followed by egress through the basement membrane into the organ parenchyma by mechanisms similar to those involved in invasion.

The site of extravasation and the organ distribution of metastases generally can be predicted by the location of the primary tumor and its vascular or lymphatic drainage. Many tumors metastasize to the organ that presents the first capillary bed they encounter after entering the circulation. *In many cases, however, the natural pathways of drainage do not readily explain the distribution of metastases.* As pointed out earlier, some tumors (e.g., lung cancers) tend to involve the adrenals quite often but almost never spread to skeletal muscle. Such organ tropism may be related to the following mechanisms:

- Expression of adhesion molecules by tumor cells whose ligands are expressed preferentially on the endothelium of target organs
- Expression of chemokines and their receptors. As discussed in Chapter 2, chemokines participate in directed movement (chemotaxis) of leukocytes, and it seems that cancer cells use similar tricks to home in on specific tissues. Human breast cancer cells express high levels of the chemokine receptors *CXCR4* and *CCR7.* The ligands for these receptors (i.e., chemokines CXCL12 and CCL21) are highly expressed only in those organs to which breast cancer cells metastasize. On the basis of this observation, it is speculated that blockade of chemokine receptors may limit metastases.

- Once they reach a target, the tumor cells must be able to colonize the site. Factors that regulate colonization are not completely understood. However, it is known that after extravasation, tumor cells are dependent on a receptive stroma for growth. Thus, in some cases, the target tissue may be a nonpermissive environment—unfavorable soil, so to speak, for the growth of tumor seedlings. For example, although well vascularized, skeletal muscles are rarely the site of metastases.

Despite their "cleverness" in escaping their sites of origin, tumor cells are quite inefficient in colonizing of distant organs. Millions of tumor cells are shed daily from even small tumors. These cells can be detected in the bloodstream and in small foci in the bone marrow, even in patients in whom gross metastatic lesions never develop. Indeed, the concept of dormancy, referring to the prolonged survival of micrometastases without progression, is well described in melanoma and in breast and prostate cancer.

Although the molecular mechanisms of colonization are just beginning to be unraveled in mouse models, a consistent theme seems to be that tumor cells secrete cytokines, growth factors, and proteases that act on the resident stromal cells, which in turn make the metastatic site habitable for the cancer cell. With a better molecular understanding of the mechanisms of metastasis, the clinician's ability to target them therapeutically will be greatly enhanced. Despite the foregoing considerations, the precise localization of metastases cannot be predicted with any form of cancer. Evidently, many tumors have not read the relevant chapters of the pathology textbooks!

Molecular Genetics of Metastasis

A long-held theory of tumor progression suggests that as tumors grow, individual cells randomly accumulate mutations, creating subclones with distinct combinations of mutations. According to this hypothesis, only a small subpopulation of the tumor cells contains all of the mutations necessary for metastasis. Recent experiments, however, in which gene profiling was performed for primary tumors and for metastatic deposits, have challenged this hypothesis. For example, a subset of breast cancers has a gene expression signature similar to that found in metastases, although no clinical evidence for metastasis is apparent. In these tumors, most if not all cells apparently acquire a predilection for metastatic spread early on, during primary carcinogenesis. Metastasis, according to this view, is not dependent on the stochastic generation of metastatic subclones during tumor progression, but is an intrinsic property of the tumor developed during carcinogenesis. Of note, however, gene expression analyses like those just described would not detect a small subset of metastatic subclones within a large tumor. Perhaps both mechanisms are operative, with aggressive tumors acquiring a metastasis-permissive gene expression pattern early in tumorigenesis that requires some additional random mutations to complete the metastatic phenotype.

An open question in cancer biology is whether there are genes whose principal or sole contribution to tumorigenesis is to control metastases. This question is of more than academic interest, because if altered forms of certain genes promote or suppress the metastatic phenotype, their detection in a primary tumor would have both prognostic and therapeutic implications. Among candidates for such metastasis oncogenes are those encoding SNAIL and TWIST, transcription factors whose primary function is to promote epithelial-to-mesenchymal transition (EMT). In EMT, carcinoma cells downregulate certain epithelial markers (e.g., E-cadherin) and upregulate certain mesenchymal markers (e.g., vimentin, smooth muscle actin). These molecular changes are accompanied by phenotypic alterations such as morphologic change from polygonal epithelioid cell shape to a spindly mesenchymal shape, along with increased production of proteolytic enzymes that promote migration and invasion. These changes are believed to favor the development of a promigratory phenotype that is essential for metastasis. Loss of E-cadherin expression seems to be a key event in EMT, and SNAIL and TWIST are transcriptional repressors that promote EMT by downregulating E-cadherin expression. How expression of these master regulator transcription factors is stimulated in tumors is not clear; however, experimental models suggest that interactions of tumor cells with stromal cells are a key stimulus for this change. Thus, acquisition of a metastatic phenotype may not require a set of mutations but may be an emergent property arising from the interactions of tumor cells and stroma.

SUMMARY

Invasion and Metastasis

- Ability to invade tissues, a hallmark of malignancy, occurs in four steps: loosening of cell–cell contacts, degradation of ECM, attachment to novel ECM components, and migration of tumor cells.
- Cell–cell contacts are lost by the inactivation of E-cadherin through a variety of pathways.
- Basement membrane and interstitial matrix degradation is mediated by proteolytic enzymes secreted by tumor cells and stromal cells, such as MMPs and cathepsins.
- Proteolytic enzymes also release growth factors sequestered in the ECM and generate chemotactic and angiogenic fragments from cleavage of ECM glycoproteins.
- The metastatic site of many tumors can be predicted by the location of the primary tumor. Many tumors arrest in the first capillary bed they encounter (lung and liver, most commonly).
- Some tumors show organ tropism, probably due to activation of adhesion or chemokine receptors whose ligands are expressed by endothelial cells at the metastatic site.

Reprogramming Energy Metabolism

Reprogramming of energy metabolism is so common to tumors that it is now considered a hallmark of cancer. Even in the presence of ample oxygen, cancer cells shift their glucose metabolism away from the oxygen-hungry but efficient mitochondria to glycolysis. This phenomenon, called

the Warburg effect and also known as aerobic glycolysis, has been recognized for many years (indeed, Otto Warburg received the Nobel prize for discovery of the effect that bears his name in 1931) but was largely neglected until recently.

As is well known, aerobic glycolysis is less efficient than mitochondrial oxidative phosphorylation, producing 2 molecules of ATP per molecule of glucose, versus 36. Yet tumors that adopt aerobic glycolysis, such as Burkitt lymphoma, are the most rapidly growing of human cancers. Indeed, in clinical practice, the "glucose hunger" of such tumors is used to visualize tumors by positron emission tomography (PET) scanning, in which the patient is injected with ^{18}F-fluorodeoxyglucose, a nonmetabolizable derivative of glucose. Most tumors are PET-positive, and rapidly growing ones are markedly so.

Importantly, it is now recognized that rapidly dividing normal cells, such as those in the embryo, also adopt Warburg metabolism, indicating that this mode of metabolism is favored when rapid growth is required. How can this be, given that aerobic glycolysis generates much less ATP per mole of glucose? In addition to doubling its DNA content before division, an actively dividing cell (whether normal or transformed) must also double all of its other components, including membranes, proteins, and organelles. This task requires increased uptake of nutrients, particularly glucose and amino acids. Studies of intermediate metabolism suggest that in rapidly growing cells glucose is the primary source of the carbons that are used for synthesis of lipids (needed for membrane assembly) as well as other metabolites needed for nucleic acid synthesis. This pattern of glucose carbon use is achieved by shunting pyruvate toward biosynthetic pathways at the expense of the oxidative phosphorylation pathway and ATP generation. Thus, the metabolism of cancer can also be viewed from a darwinian perspective; tumor cells that adapt this altered metabolism are able to divide more rapidly and outpace competing tumor cells that do not.

Since aerobic glycolysis continues in tumors in the face of adequate oxygen, it follows that the changes that promote the switch in metabolism must have become hard wired in the tumor cell. It is now becoming clear that oncogenes and tumor suppressors that favor cell growth, such as TP53, PTEN, and Akt (an intermediary in RAS signaling) stimulate glucose uptake by affecting glucose transporter proteins and favor aerobic glycolysis. Indeed the Warburg effect appears to be sufficiently central to the cancer phenotype that drugs that target this pathway are being developed for therapy.

Evasion of the Immune System

As mentioned at the outset, the ability of tumors to evade destruction by the immune system (like the reprogramming of the energy metabolism) is now considered a hallmark of cancer. Most tumors arise in immunocompetent hosts; accordingly, a likely strategy for success is to trick the immune system in such a way that the tumor fails to be recognized or eliminated despite the fact the affected person's body has an army of cells that are quite capable of thwarting a microbial infection or rejecting an allogeneic organ transplant. Discussion of this hallmark is postponed to a later section, since it is best understood in the context of the nature of tumor antigens and how they might be recognized.

Genomic Instability as an Enabler of Malignancy

The preceding section identified eight defining features of malignancy and the genetic alterations that are responsible for the phenotypic attributes of cancer cells. How do these mutations arise? Although humans are awash in environmental agents that are mutagenic (e.g., chemicals, radiation, sunlight), cancers are relatively rare outcomes of these encounters. This state of affairs results from the ability of normal cells to repair DNA damage. The importance of DNA repair in maintaining the integrity of the genome is highlighted by several inherited disorders in which genes that encode proteins involved in DNA repair are defective. *Persons born with such inherited defects in DNA repair proteins are at greatly increased risk for the development of cancer.* Typically, genomic instability occurs when both copies of the gene are lost; however, recent work has suggested that at least a subset of these genes may promote cancer in a haploinsufficient manner. Defects in three types of DNA repair systems—mismatch repair, nucleotide excision repair, and recombination repair—are presented next. While these discussions focus on inherited syndromes, a point worthy of emphasis is that sporadic cancers often incur mutations in these genes as well, which in turn enable the accumulation of mutations in other genes whose dysfunction contributes to the hallmarks of cancer.

Hereditary Nonpolyposis Colon Cancer Syndrome

The role of DNA repair genes in predisposition to cancer is illustrated dramatically by hereditary nonpolyposis colon carcinoma (HNPCC) syndrome. This disorder, characterized by familial carcinomas of the colon affecting predominantly the cecum and proximal colon (Chapter 14), results from defects in genes involved in DNA mismatch repair. When a strand of DNA is being repaired, these genes act as "spell checkers." For example, if there is an erroneous pairing of G with T, rather than the normal A with T, the mismatch repair genes correct the defect. Without these "proofreaders," errors accumulate at an increased rate, a so-called mutator phenotype. Mutations in at least four mismatch repair genes have been found to underlie HNPCC (Chapter 14). Each affected person inherits one defective copy of one of several DNA mismatch repair genes and acquires the second hit in colonic epithelial cells. Thus, DNA repair genes affect cell growth only indirectly—by allowing mutations in other genes during the process of normal cell division. A characteristic finding in the genome of patients with mismatch repair defects is microsatellite instability (MSI). Microsatellites are tandem repeats of one to six nucleotides found throughout the genome. In normal people, the length of these microsatellites remains constant. By contrast, in patients with HNPCC, these satellites are unstable and increase or decrease in length. Although HNPCC accounts for only 2% to 4% of all colonic cancers, MSI can be detected in about 15% of sporadic cancers. The growth-regulating genes that are

mutated in HNPCC include those encoding TGF-β receptor type II, BAX, and other oncogenes and tumor suppressor genes.

Xeroderma Pigmentosum

Patients with another inherited disorder, xeroderma pigmentosum, are at increased risk for the development of cancers of sun-exposed skin. The basis for this disorder is defective DNA repair. Ultraviolet (UV) rays in sunlight cause cross-linking of pyrimidine residues, preventing normal DNA replication. Such DNA damage is repaired by the nucleotide excision repair system. Several proteins are involved in nucleotide excision repair, and an inherited loss of any one of these can give rise to xeroderma pigmentosum.

Diseases with Defects in DNA Repair by Homologous Recombination

A group of autosomal recessive disorders comprising Bloom syndrome, ataxia-telangiectasia, and Fanconi anemia is characterized by hypersensitivity to other DNA-damaging agents, such as ionizing radiation (in Bloom syndrome and ataxia-telangiectasia), or to DNA cross-linking agents, such as nitrogen mustard (in Fanconi anemia). Their phenotype is complex and includes, in addition to predisposition to cancer, features such as neural symptoms (in ataxia-telangiectasia), anemia (in Fanconi anemia), and developmental defects (in Bloom syndrome). The gene mutated in ataxia-telangiectasia is *ATM,* which encodes a protein kinase that is important in recognizing DNA damage caused by ionizing radiation and initiating p53 activation.

Evidence for the role of DNA repair genes in the origin of cancer also comes from the study of hereditary breast cancer. Mutations in two genes, *BRCA1* and *BRCA2,* account for 50% of cases of familial breast cancer. In addition to breast cancer, women with *BRCA1* mutations have a substantially higher risk of epithelial ovarian cancers, and men have a slightly higher risk of prostate cancer. Likewise, mutations in the *BRCA2* gene increase the risk of breast cancer in both men and women, as well as cancer of the ovary, prostate, pancreas, bile ducts, stomach, melanocytes, and B lymphocytes. Although the functions of these genes have not been elucidated fully, cells that lack these genes develop chromosomal breaks and severe aneuploidy. Indeed, both genes seem to function, at least in part, in the homologous recombination DNA repair pathway. For example, BRCA1 forms a complex with other proteins in the homologous recombination pathway and also is linked to the ATM kinase pathway. *BRCA2* was identified as one of several genes mutated in Fanconi anemia, and the BRCA2 protein has been shown to bind to RAD51, a protein required for homologous recombination. Similar to other tumor suppressor genes, both copies of *BRCA1* and *BRCA2* must be inactivated for cancer to develop. Although linkage of *BRCA1* and *BRCA2* to familial breast cancers is established, these genes are rarely inactivated in sporadic cases of breast cancer. In this regard, *BRCA1* and *BRCA2* are different from other tumor suppressor genes, such as *APC* and *TP53,* which are inactivated in both familial and sporadic cancers.

Cancers Resulting From Mutations Induced by Regulated Genomic Instability: Lymphoid Neoplasms

A special type of DNA damage plays a central role in the pathogenesis of tumors of B and T lymphocytes. As described earlier, adaptive immunity relies on the ability of B and T cells to diversify their antigen receptor genes. Early B and T cells both express a pair of gene products, RAG1 and RAG2, that carry out V(D)J segment recombination, permitting the assembling of functional antigen receptor genes. In addition, after encountering antigen, mature B cells express a specialized enzyme called activation-induced cytosine deaminase (AID), which catalyzes both immunoglobulin gene class switch recombination and somatic hypermutation. Errors during antigen receptor gene assembly and diversification are responsible for many of the mutations that cause lymphoid neoplasms, described in detail in Chapter 11.

SUMMARY

Genomic Instability as Enabler of Malignancy

- Persons with inherited mutations of genes involved in DNA repair systems are at greatly increased risk for the development of cancer.
- Patients with HNPCC syndrome have defects in the mismatch repair system, leading to development of carcinomas of the colon. These patients' genomes show MSI, characterized by changes in length of short tandem repeating sequences throughout the genome.
- Patients with xeroderma pigmentosum have a defect in the nucleotide excision repair pathway and are at increased risk for the development of cancers of the skin exposed to UV light, because of an inability to repair pyrimidine dimers.
- Syndromes involving defects in the homologous recombination DNA repair system constitute a group of disorders—Bloom syndrome, ataxia-telangiectasia, and Fanconi anemia—that are characterized by hypersensitivity to DNA-damaging agents, such as ionizing radiation. *BRCA1* and *BRCA2*, which are mutated in familial breast cancers, are involved in DNA repair.
- Mutations incurred in lymphoid cells expressing gene products that induce genomic instability (RAG1, RAG2, AID) are important causes of lymphoid neoplasms.

Tumor-Promoting Inflammation as Enabler of Malignancy

Accumulating evidence suggests that inflammation, often thought of as a protective response against tumors, can paradoxically also enable malignancy. This occurs in two different settings:

- *Persistent chronic inflammation in response to microbial infections or as part of an autoimmune reaction*. This is exemplified by the increased risk of cancer in patients affected by a variety of chronic inflammatory diseases of the gastrointestinal tract. These include Barrett esophagus, ulcerative colitis, *H. pylori* gastritis, hepatitis B and C, and chronic pancreatitis. As with any cause of chronic tissue injury, there is a compensatory proliferation of

cells in an attempt to repair the damage. This regenerative process is aided and abetted by a plethora of growth factors, cytokines, chemokines, and other bioactive substances produced by activated immune cells collected at the site. Persistent cell replication and reduced apoptosis under these conditions place the cells at risk of acquiring mutations in one or more of the genes involved in carcinogenesis. In addition, inflammatory cells such as neutrophils can contribute to carcinogenesis by secretion of reactive oxygen species, which in turn can inflict additional DNA damage in rapidly dividing cells.

- *When inflammation occurs in response to tumors.* Pathologists have known for quite some time that many tumors are infiltrated by leukocytes. The degree of inflammation varies, but virtually every tumor contains cells of the adaptive and innate components of the immune system. The conventional wisdom has been that the inflammatory reaction is protective since it represents an attempt by the host to destroy the tumor. Indeed, that may well be the purpose of the inflammatory reaction, but these cells can exert tumor-promoting activity by producing growth factors and inflicting additional DNA damage as described above.

Whatever the precise mechanism, the link between inflammation and cancer has practical implications. For instance, expression of the enzyme cyclooxygenase-2 (COX-2), which brings about the conversion of arachidonic acid into prostaglandins (Chapter 2), is induced by inflammatory stimuli and is increased in colon cancers and other tumors. The use of COX-2 inhibitors for cancer prevention and treatment is an active area of research.

Important clinical considerations emerge from the principles presented in the foregoing discussion of the hallmarks of cancer: These hallmarks provide a road map for the development of new therapeutic agents for the treatment of cancer (Fig. 5–29).

Multistep Carcinogenesis and Cancer Progression

As described earlier, the acquisition of several fundamental abnormalities is a prerequisite to development of malignancy. It follows, then, that *each cancer must result from accumulation of multiple mutations.* A dramatic example of incremental acquisition of the malignant phenotype is documented by the study of colon carcinoma. These lesions are believed to evolve through a series of morphologically identifiable stages: colon epithelial hyperplasia followed by formation of adenomas that progressively enlarge and ultimately undergo malignant transformation (Chapter 14). The proposed molecular correlates of this adenoma-carcinoma sequence are illustrated in Figure 5–30. According to this scheme, inactivation of the *APC* tumor suppressor gene occurs first, followed by activation of *RAS* and, ultimately, loss of a tumor suppressor gene on 18q and loss of *TP53*. The precise temporal sequence of mutations may be different in different tumors.

ETIOLOGY OF CANCER: CARCINOGENIC AGENTS

Genetic damage lies at the heart of carcinogenesis. What extrinsic agents can inflict such damage? Three classes of carcinogenic agents have been identified: (1) chemicals, (2) radiant energy, and (3) microbial agents. Chemicals and radiant energy are documented causes of cancer in humans, and oncogenic viruses are involved in the pathogenesis of tumors in several animal models and some human tumors. In the following discussion, each class of agent is considered separately; of note, however, several may act in concert or sequentially to produce the multiple genetic abnormalities characteristic of neoplastic cells.

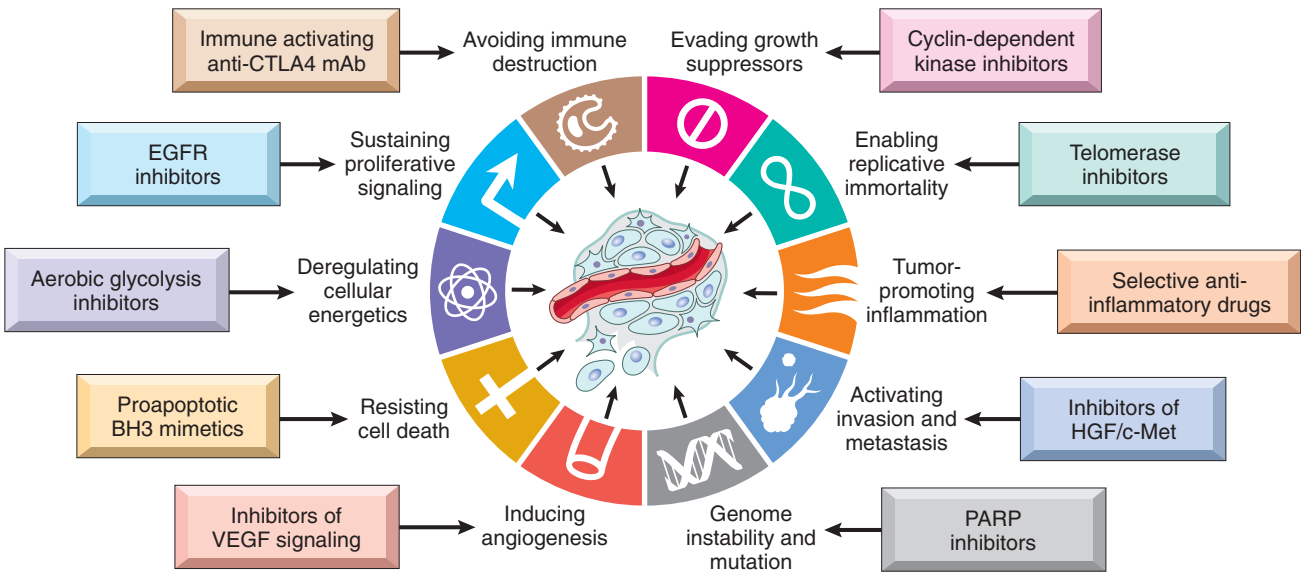


Figure 5–29 Therapeutic targeting of hallmarks of cancer.
(From Hanahan D, Weiberg RA: The hallmarks of cancer: the next generation. Cell 144:646, 2011.)

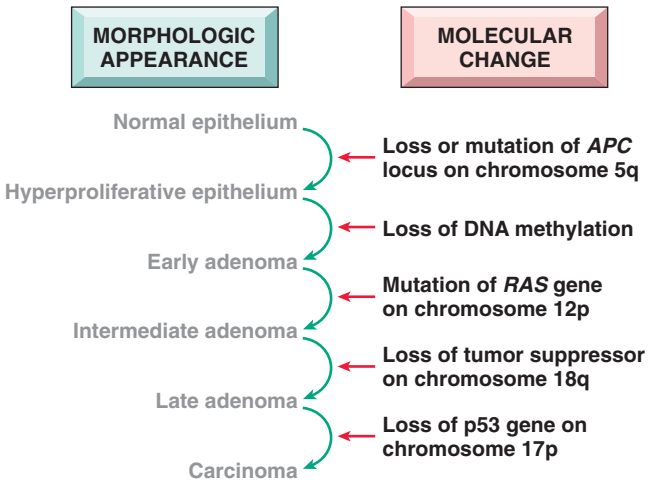


Figure 5–30 Molecular model for the evolution of colorectal cancers through the adenoma–carcinoma sequence.
(Data from studies by Fearon ER, Vogelstein B: A genetic model of colorectal carcinogenesis. Cell 61:759, 1990.)

Chemical Carcinogens

More than 200 years ago, the London surgeon Sir Percival Pott correctly attributed scrotal skin cancer in chimney sweeps to chronic exposure to soot. On the basis of this observation, the Danish Chimney Sweeps Guild ruled that its members must bathe daily. No public health measure since that time has achieved so much in the control of a form of cancer. Subsequently, hundreds of chemicals have been shown to be carcinogenic in animals.

Some of the major agents are presented in Table 5–4. A few comments on a handful of these are offered next.

Direct-Acting Agents

Direct-acting agents require no metabolic conversion to become carcinogenic. They are in general weak carcinogens but are important because some of them are cancer chemotherapy drugs (e.g., alkylating agents) used in regimens that may cure certain types of cancer (e.g., Hodgkin lymphoma), only to evoke a subsequent, second form of cancer, usually leukemia. This situation is even more tragic when the initial use of such agents has been for nonneoplastic disorders, such as rheumatoid arthritis or Wegener granulomatosis. The associated risk of induced cancer is low, but its existence dictates judicious use of such agents.

Indirect-Acting Agents

The designation *indirect-acting* refers to chemicals that require metabolic conversion to an *ultimate carcinogen*. Some of the most potent indirect chemical carcinogens are polycyclic hydrocarbons, present in fossil fuels. For example, benzo[*a*]pyrene and other carcinogens are formed in the high-temperature combustion of tobacco in cigarette smoking. *These products are implicated in the causation of lung cancer in cigarette smokers.* Polycyclic hydrocarbons also may be produced from animal fats during the process of broiling meats and are present in smoked meats and fish.

Table 5–4 Major Chemical Carcinogens

Direct-Acting Carcinogens
Alkylating Agents
β-Propiolactone
Dimethyl sulfate
Diepoxybutane
Anticancer drugs (cyclophosphamide, chlorambucil, nitrosoureas, and others)
Acylating Agents
1-Acetyl-imidazole
Dimethylcarbamyl chloride
Procarcinogens That Require Metabolic Activation
Polycyclic and Heterocyclic Aromatic Hydrocarbons
Benz(*a*)anthracene
Benzo(*a*)pyrene
Dibenz(*a*,*h*)anthracene
3-Methylcholanthrene
7, 12-Dimethylbenz(*a*)anthracene
Aromatic Amines, Amides, Azo Dyes
2-Naphthylamine (β-naphthylamine)
Benzidine
2-Acetylaminofluorene
Dimethylaminoazobenzene (butter yellow)
Natural Plant and Microbial Products
Aflatoxin B_1
Griseofulvin
Cycasin
Safrole
Betel nuts
Others
Nitrosamine and amides
Vinyl chloride, nickel, chromium
Insecticides, fungicides
Polychlorinated biphenyls

The principal active products in many hydrocarbons are epoxides, which form covalent adducts (addition products) with molecules in the cell, principally DNA, but also with RNA and proteins.

The aromatic amines and azo dyes constitute another class of indirect-acting carcinogens. Before its carcinogenicity was recognized, β-naphthylamine was responsible for a 50-fold increased incidence of bladder cancers in heavily exposed workers in the aniline dye and rubber industries. Many other occupational carcinogens are listed in Table 5–2. Because indirect-acting carcinogens require metabolic activation for their conversion to DNA-damaging agents, much interest is focused on the enzymatic pathways that are involved, such as that mediated by the cytochrome P-450–dependent monooxygenases. The genes that encode these enzymes are polymorphic, and enzyme activity varies among different persons. It is widely believed that the susceptibility to chemical carcinogenesis depends at least in part on the specific allelic form of the enzyme inherited. Thus, it may be possible in the future to assess cancer risk in a given patient by genetic analysis of such enzyme polymorphisms.

A few other agents merit brief mention. Aflatoxin B_1 is of interest because it is a naturally occurring agent produced by some strains of *Aspergillus,* a mold that grows on improperly stored grains and nuts. A strong correlation has been found between the dietary level of this food contaminant and the incidence of hepatocellular carcinoma in some parts of Africa and the Far East. Additionally, vinyl chloride, arsenic, nickel, chromium, insecticides, fungicides, and polychlorinated biphenyls are potential carcinogens in the workplace and about the house. Finally, nitrites used as food preservatives have caused concern, since they cause nitrosylation of amines contained in the food. The nitrosamines thus formed are suspected to be carcinogenic.

Mechanisms of Action of Chemical Carcinogens

Because malignant transformation results from mutations, it should come as no surprise that most chemical carcinogens are mutagenic. Indeed, all direct and ultimate carcinogens contain highly reactive electrophile groups that form chemical adducts with DNA, as well as with proteins and RNA. Although any gene may be the target of chemical carcinogens, the commonly mutated oncogenes and tumor suppressors, such as *RAS* and *TP53,* are important targets of chemical carcinogens. Indeed, specific chemical carcinogens, such as aflatoxin B_1, produce characteristic mutations in the *TP53* gene, such that detection of the "*signature mutation*" within the *TP53* gene establishes aflatoxin as the causative agent. These associations are proving to be useful tools in epidemiologic studies of chemical carcinogenesis.

Carcinogenicity of some chemicals is augmented by subsequent administration of *promoters* (e.g., phorbol esters, hormones, phenols, certain drugs) that by themselves are nontumorigenic. To be effective, repeated or sustained exposure to the promoter must *follow* the application of the mutagenic chemical, or *initiator.* The initiation-promotion sequence of chemical carcinogenesis raises an important question: Since promoters are not mutagenic, how do they contribute to tumorigenesis? Although the effects of tumor promoters are pleiotropic, *induction of cell proliferation is a sine qua non of tumor promotion.* It seems most likely that while the application of an initiator may cause the mutational activation of an oncogene such as *RAS,* subsequent application of promoters leads to clonal expansion of initiated (mutated) cells. Forced to proliferate, the initiated clone of cells accumulates additional mutations, developing eventually into a malignant tumor. Indeed, the concept that sustained cell proliferation increases the risk of mutagenesis, and hence promotes neoplastic transformation, also is applicable to human carcinogenesis. For example, endometrial hyperplasia (Chapter 18) and increased regenerative activity that accompanies chronic liver cell injury are associated with the development of cancer in these organs. Were it not for the DNA repair mechanisms discussed earlier, the incidence of chemically induced cancers in all likelihood would be much higher. As mentioned previously, the rare hereditary disorders of DNA repair, including xeroderma pigmentosum, are associated with greatly increased risk of cancers induced by UV light and certain chemicals.

SUMMARY

Chemical Carcinogens

- Chemical carcinogens have highly reactive electrophile groups that directly damage DNA, leading to mutations and eventually cancer.
- Direct-acting agents do not require metabolic conversion to become carcinogenic, while indirect-acting agents are not active until converted to an ultimate carcinogen by endogenous metabolic pathways. Hence, polymorphisms of endogenous enzymes such as cytochrome P-450 may influence carcinogenesis.
- After exposure of a cell to a mutagen or an initiator, tumorigenesis can be enhanced by exposure to promoters, which stimulate proliferation of the mutated cells.
- Examples of human carcinogens are direct-acting agents (e.g., alkylating agents used for chemotherapy), indirect-acting agents (e.g., benzopyrene, azo dyes, aflatoxin), and promoters or agents that cause hyperplasia of endometrium or regenerative activity in the liver.

Radiation Carcinogenesis

Radiation, whatever its source (UV rays of sunlight, x-rays, nuclear fission, radionuclides) is an established carcinogen. Unprotected miners of radioactive elements have a 10-fold increased incidence of lung cancers. Follow-up study of survivors of the atomic bombs dropped on Hiroshima and Nagasaki disclosed a markedly increased incidence of leukemia—principally myelogenous leukemias—after an average latent period of about 7 years, as well as increased mortality rates for thyroid, breast, colon, and lung carcinomas. The nuclear power accident at Chernobyl in the former Soviet Union continues to exact its toll in the form of high cancer incidence in the surrounding areas. More recently, it is feared that radiation release from a nuclear power plant in Japan damaged by a massive earthquake and tsunami will result in significantly increased cancer incidence in the surrounding geographic areas.

Therapeutic irradiation of the head and neck can give rise to papillary thyroid cancers years later. The oncogenic properties of ionizing radiation are related to its mutagenic effects; it causes chromosome breakage, translocations, and, less frequently, point mutations. Biologically, double-stranded DNA breaks seem to be the most important form of DNA damage caused by radiation.

The oncogenic effect of UV rays merits special mention because it highlights the importance of DNA repair in carcinogenesis. Natural UV radiation derived from the sun can cause skin cancers (melanomas, squamous cell carcinomas, and basal cell carcinomas). At greatest risk are fair-skinned people who live in locales such as Australia and New Zealand that receive a great deal of sunlight. Nonmelanoma skin cancers are associated with total cumulative exposure to UV radiation, whereas melanomas are associated with intense intermittent exposure—as occurs with sunbathing. UV light has several biologic effects on cells. Of particular relevance to carcinogenesis is the ability to damage DNA by forming pyrimidine dimers.

This type of DNA damage is repaired by the nucleotide excision repair pathway. With extensive exposure to UV light, the repair systems may be overwhelmed, and skin cancer results. As mentioned earlier, patients with the inherited disease *xeroderma pigmentosum* have a defect in the nucleotide excision repair pathway. As expected, there is a greatly increased predisposition to skin cancers in this disorder.

SUMMARY

Radiation Carcinogenesis

- Ionizing radiation causes chromosome breakage, translocations, and, less frequently, point mutations, leading to genetic damage and carcinogenesis.
- UV rays induce the formation of pyrimidine dimers within DNA, leading to mutations. Therefore, UV rays can give rise to squamous cell carcinomas and melanomas of the skin.

Viral and Microbial Oncogenesis

Many DNA and RNA viruses have proved to be oncogenic in animals as disparate as frogs and primates. Despite intense scrutiny, however, only a few viruses have been linked with human cancer. The following discussion focuses on human oncogenic viruses. Also discussed is the emerging role of the bacterium *H. pylori* in gastric cancer.

Oncogenic RNA Viruses

The study of oncogenic retroviruses in animals has provided spectacular insights into the genetic basis of cancer. However, only one retrovirus, the human T cell lymphotropic virus-1 (HTLV-1), has been demonstrated to cause cancer in humans. HTLV-1 is associated with a form of T cell leukemia/lymphoma that is endemic in certain parts of Japan and the Caribbean basin but is found sporadically elsewhere, including the United States. Similar to the human immunodeficiency virus (HIV), HTLV-1 has tropism for CD4+ T cells, and this subset of T cells is the major target for neoplastic transformation. Human infection requires transmission of infected T cells through sexual intercourse, blood products, or breastfeeding. Leukemia develops only in about 3% to 5% of infected persons after a long latent period of 20 to 50 years.

There is little doubt that HTLV-1 infection of T lymphocytes is necessary for leukemogenesis, but the molecular mechanisms of transformation are not clear. The HTLV-1 genome does not contain a viral oncogene, and in contrast with certain animal retroviruses, no consistent integration site next to a cellular oncogene has been discovered. Indeed, the long latency period between initial infection and development of disease suggests a multistep process, during which many oncogenic mutations are accumulated.

The genome of HTLV-1 contains, in addition to the usual retroviral genes, a unique region called *pX*. This region contains several genes, including one called *TAX*. The TAX protein has been shown to be necessary and sufficient for cellular transformation. By interacting with several transcription factors, such as NF-κB, the TAX protein can transactivate the expression of genes that encode cytokines, cytokine receptors, and costimulatory molecules. This inappropriate gene expression leads to autocrine signaling loops and increased activation of promitogenic signaling cascades. Furthermore, TAX can drive progression through the cell cycle by directly binding to and activating cyclins. In addition, TAX can repress the function of several tumor suppressor genes that control the cell cycle, including *CDKN2A/p16* and *TP53*. From these and other observations, the following scenario is emerging (Fig. 5–31): The *TAX* gene turns on several cytokine genes and their receptors (e.g., the interleukins IL-2 and IL-2R and IL-15 and IL-15R), setting up an autocrine system that drives T cell proliferation. Of these cytokines, IL-15 seems to be more important, but much remains to be defined. Additionally, a parallel paracrine pathway is activated by increased production of granulocyte-macrophage colony-stimulating factor, which stimulates neighboring macrophages to produce other T cell mitogens. Initially, the T cell proliferation is polyclonal, because the virus infects many cells, but because of TAX-based inactivation of tumor suppressor genes such as *TP53*, the proliferating T cells are at increased risk for secondary transforming events (mutations), which lead ultimately to the outgrowth of a monoclonal neoplastic T cell population.

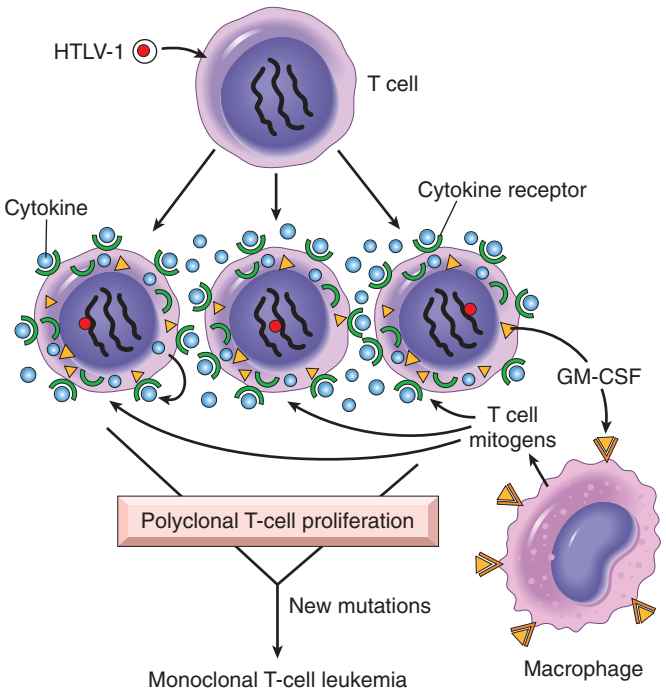


Figure 5–31 Pathogenesis of human T cell lymphotropic virus (HTLV-1)–induced T cell leukemia/lymphoma. HTLV-1 infects many T cells and initially causes polyclonal proliferation by autocrine and paracrine pathways triggered by the *TAX* gene. Simultaneously, TAX neutralizes growth inhibitory signals by affecting *TP53* and *CDKN2A/p16* genes. Ultimately, a monoclonal T cell leukemia/lymphoma results when one proliferating T cell suffers additional mutations.

SUMMARY

Oncogenic RNA Viruses

- HTLV-I causes a T cell leukemia that is endemic in Japan and the Caribbean.
- The HTLV-I genome encodes a viral TAX protein, which turns on genes for cytokines and their receptors in infected T cells. This sets up autocrine and paracrine signaling loops that stimulate T cell proliferation. Although this proliferation initially is polyclonal, the proliferating T cells are at increased risk for secondary mutations that lead to the outgrowth of a monoclonal leukemia.

Oncogenic DNA Viruses

As with RNA viruses, several oncogenic DNA viruses that cause tumors in animals have been identified. Four DNA viruses—HPV, Epstein-Barr virus (EBV), Kaposi sarcoma herpesvirus (KSHV, also called human herpesvirus-8 [HHV-8]), and hepatitis B virus (HBV)—are of special interest because they are strongly associated with human cancer. KSHV and Kaposi sarcoma are discussed in Chapter 4. The others are presented here.

Human Papillomavirus

Scores of genetically distinct types of HPV have been identified. Some types (e.g., 1, 2, 4, and 7) cause benign squamous papillomas (warts) in humans (Chapters 18 and 21). Genital warts have low malignant potential and are also associated with low-risk HPVs, predominantly HPV-6 and HPV-11. By contrast, high-risk HPVs (e.g., types 16 and 18) cause several cancers, particularly squamous cell carcinoma of the cervix and anogenital region. In addition, at least 20% of oropharyngeal cancers, particularly those arising in the tonsils, are associated with HPV.

The oncogenic potential of HPV can be related to products of two early viral genes, E6 and E7. Together, they interact with a variety of growth-regulating proteins encoded by proto-oncogenes and tumor suppressor genes. The E7 protein binds to the retinoblastoma protein and releases the E2F transcription factors that normally are sequestered by Rb, promoting progression through the cell cycle. Of interest, E7 protein from high-risk HPV types has a higher affinity for Rb than does E7 from low-risk HPV types. E7 also inactivates the CDKIs CDKN1A/p21 and CDNK1B/p27. The E6 protein has complementary effects. It binds to and mediates the degradation of p53. By analogy with E7, E6 from high-risk HPV types has a higher affinity for p53 than does E6 from low-risk HPV types. Also of interest, in benign warts the HPV genome is maintained in a nonintegrated episomal form, while in cancers the HPV genome is randomly integrated into the host genome. Integration interrupts the viral DNA, resulting in overexpression of the oncoproteins E6 and E7. Furthermore, cells in which the viral genome has integrated show significantly more genomic instability.

To summarize, infection with high-risk HPV types simulates the loss of tumor suppressor genes, activates cyclins, inhibits apoptosis, and combats cellular senescence. Thus, it is evident that many of the hallmarks of cancer discussed earlier are driven by HPV proteins. However, infection with HPV itself is not sufficient for carcinogenesis. For example, when human keratinocytes are transfected with DNA from HPV-16, -18, or -31 in vitro, they are immortalized, but they do not form tumors in experimental animals. Cotransfection with a mutated *RAS* gene results in full malignant transformation. These data strongly suggest that HPV, in all likelihood, acts in concert with other environmental factors (Chapter 18). However, the primacy of HPV infection in the causation of cervical cancer is attested to by the near-complete protection from this cancer by anti-HPV vaccines.

Epstein-Barr Virus

EBV was the first virus linked to a human tumor, Burkitt lymphoma. Over the last 40 years, however, EBV has been discovered with the cells of a surprisingly diverse list of tumors, including B cell lymphomas in patients with defective T cell immunity (e.g., those infected with HIV), a subset of Hodgkin lymphoma, nasopharyngeal carcinoma, a subset of T cell lymphomas, gastric carcinomas, NK cell lymphomas, and even, in rare instances, sarcomas, mainly in the immunosuppressed.

Burkitt lymphoma is endemic in certain parts of Africa and is sporadic elsewhere. In endemic areas, tumor cells in virtually all affected patients carry the EBV genome. The molecular basis for B cell proliferations induced by EBV is complex. EBV uses the complement receptor CD21 to attach to and infect B cells. In vitro, such infection leads to polyclonal B cell proliferation and generation of B lymphoblastoid cell lines. One of the EBV-encoded genes, called *LMP1* (latent membrane protein 1) acts as an oncogene, and its expression in transgenic mice induces B cell lymphomas. LMP1 promotes B cell proliferation by activating signaling pathways, such as NF-κB and JAK/STAT, which mimic B cell activation by the B cell surface molecule CD40. Concurrently, LMP1 prevents apoptosis by activating BCL2. Thus, the virus "borrows" a normal B cell activation pathway to promote its own replication by expanding the pool of cells susceptible to infection. Another EBV-encoded protein, EBNA2, transactivates several host genes, including cyclin D and the *src* family of proto-oncogenes. In addition, the EBV genome contains a viral cytokine, vIL-10, that was pirated from the host genome. This viral cytokine can prevent macrophages and monocytes from activating T cells and killing virally infected cells.

In immunologically normal persons, EBV-driven polyclonal B cell proliferation is readily controlled, and the affected patient either remains asymptomatic or experiences a self-limited episode of infectious mononucleosis (Chapter 11). Evasion of the immune system seems to be a key step in EBV-related oncogenesis. In regions of the world in which Burkitt lymphoma is endemic, concomitant (endemic) malaria (or other infections) impairs immune competence, allowing sustained B cell proliferation. Of interest, although *LMP1* is the primary transforming oncogene in the EBV genome, it is not expressed in EBV-associated Burkitt lymphoma, presumably because it also is one of the major viral antigens recognized by the immune system. Infected cells expressing viral antigens such as LMP-1 are kept in check by the immune system. Lymphoma cells may emerge only when translocations activate the *MYC* oncogene, a consistent feature of this tumor. MYC may substitute for LMP1 signaling, allowing the tumor

cells to downregulate LMP1 and evade the immune system. Of note, in nonendemic areas, 80% of tumors are negative for EBV, but virtually all tumors possess *MYC* translocations. This observation suggests that although non-African Burkitt lymphomas are triggered by mechanisms other than EBV, these cancers develop by similar pathways.

In patients with deficient T cell function, including those with HIV and organ transplant recipients, EBV-infected B cells undergo polyclonal expansion, producing lymphoblastoid-like cells. In contrast with Burkitt lymphoma, the B lymphoblasts in immunosuppressed patients do express viral antigens, such as LMP-1, that are recognized by T cells. These potentially lethal proliferations can be subdued if T cell immunity can be restored, as may be achieved by withdrawal of immunosuppressive drugs in transplant recipients.

Nasopharyngeal carcinoma is endemic in southern China and some other locales, and the EBV genome is found in all tumors. LMP-1 is expressed in the carcinoma cells and, as in B cells, activates the NF-κB pathway. Furthermore, LMP1 induces the expression of pro-angiogenic factors such as VEGF, FGF-2, MMP-9, and COX-2, which may contribute to oncogenesis. How EBV enters epithelial cells is unclear, as these cells fail to express the CD21 protein that serves as the EBV receptor in B cells.

SUMMARY

Oncogenic DNA Viruses

- HPV is associated with benign warts, as well as cervical cancer.
- The oncogenicity of HPV is related to the expression of two viral oncoproteins, E6 and E7; they bind to Rb and p53, respectively, neutralizing their function.
- E6 and E7 from high-risk strains of HPV (which give rise to cancers) have higher affinity for their targets than do E6 and E7 from low-risk strains of HPV (which give rise to benign warts).
- EBV is implicated in the pathogenesis of Burkitt lymphomas, lymphomas in immunosuppressed patients (HIV infection or organ transplant recipients), some forms of Hodgkin lymphoma, uncommon T cell and NK cell tumors, nasopharyngeal carcinoma, a subset of gastric carcinoma, and rarely sarcomas.
- Certain EBV gene products contribute to oncogenesis by stimulating a normal B cell proliferation pathway. Concomitant compromise of immune competence allows sustained B cell proliferation, leading eventually to development of lymphoma, with occurrence of additional mutations such as t(8;14) leading to activation of the *MYC* gene.

Hepatitis B and Hepatitis C Viruses

The epidemiologic evidence linking chronic HBV and hepatitis C virus (HCV) infection with hepatocellular carcinoma is strong (Chapter 15). It is estimated that 70% to 85% of hepatocellular carcinomas worldwide are due to infection with HBV or HCV. However, the mode of action of these viruses in tumorigenesis is not fully elucidated. The HBV and HCV genomes do not encode any viral oncoproteins, and although the HBV DNA is integrated within the human genome, there is no consistent pattern of integration in liver cells. Indeed, the oncogenic effects of HBV and HCV are multifactorial, but the dominant effect seems to be immunologically mediated chronic inflammation with hepatocyte death leading to regeneration and genomic damage. Although the immune system generally is thought to be protective, recent work has demonstrated that in the setting of unresolved chronic inflammation, as occurs in viral hepatitis or chronic gastritis caused by *H. pylori* (see further on), the immune response may become maladaptive, promoting tumorigenesis.

As with any cause of hepatocellular injury, chronic viral infection leads to the compensatory proliferation of hepatocytes. This regenerative process is aided and abetted by a plethora of growth factors, cytokines, chemokines, and other bioactive substances produced by activated immune cells that promote cell survival, tissue remodeling, and angiogenesis. The activated immune cells also produce other mediators, such as reactive oxygen species, that are genotoxic and mutagenic. A key molecular step seems to be activation of the nuclear factor-κB (NF-κB) pathway in hepatocytes caused by mediators derived from the activated immune cells. Activation of the NF-κB pathway within hepatocytes blocks apoptosis, allowing the dividing hepatocytes to incur genotoxic stress and to accumulate mutations. Although this seems to be the dominant mechanism in the pathogenesis of virus-induced hepatocellular carcinoma, both HBV and HCV also contain proteins within their genomes that may more directly promote the development of cancer. The HBV genome contains a gene known as *HBx,* and hepatocellular cancers develop in mice transgenic for this gene. *HBx* can directly or indirectly activate a variety of transcription factors and several signal transduction pathways. In addition, viral integration can cause secondary rearrangements of chromosomes, including multiple deletions that may harbor unknown tumor suppressor genes.

Although not a DNA virus, HCV also is strongly linked to the pathogenesis of liver cancer. The molecular mechanisms used by HCV are less well defined than those for HBV. In addition to chronic liver cell injury and compensatory regeneration, components of the HCV genome, such as the HCV core protein, may have a direct effect on tumorigenesis, possibly by activating a variety of growth-promoting signal transduction pathways.

SUMMARY

Hepatitis B and Hepatitis C Viruses

- Between 70% and 85% of hepatocellular carcinomas worldwide are due to infection with HBV or HCV.
- The oncogenic effects of HBV and HCV are multifactorial, but the dominant effect seems to be immunologically mediated chronic inflammation, with hepatocellular injury, stimulation of hepatocyte proliferation, and production of reactive oxygen species that can damage DNA.
- The HBx protein of HBV and the HCV core protein can activate a variety of signal transduction pathways that also may contribute to carcinogenesis.

Helicobacter pylori

First incriminated as a cause of peptic ulcers, *H. pylori* now has acquired the dubious distinction of being the first bacterium classified as a carcinogen. Indeed, *H. pylori* infection is implicated in the genesis of both gastric adenocarcinomas and gastric lymphomas.

The scenario for the development of gastric adenocarcinoma is similar to that for HBV- and HCV-induced liver cancer. It involves increased epithelial cell proliferation on a background of chronic inflammation. As in viral hepatitis, the inflammatory milieu contains numerous genotoxic agents, such as reactive oxygen species. The sequence of histopathologic changes consists of initial development of chronic inflammation/gastritis, followed by gastric atrophy, intestinal metaplasia of the lining cells, dysplasia, and cancer. This sequence takes decades to complete and occurs in only 3% of infected patients. Like those of HBV and HCV, the *H. pylori* genome also contains genes directly implicated in oncogenesis. Strains associated with gastric adenocarcinoma have been shown to contain a "pathogenicity island" that contains cytotoxin-associated A gene (*CagA*). Although *H. pylori* is noninvasive, *CagA* is injected into gastric epithelial cells, where it has a variety of effects, including the initiation of a signaling cascade that mimics unregulated growth factor stimulation.

As mentioned previously, *H. pylori* is associated with an increased risk for the development of gastric lymphomas as well. The gastric lymphomas are of B cell origin, and because the transformed B cells grow in a pattern resembling that of normal mucosa-associated lymphoid tissue (MALT), they also have been referred to as MALT lymphomas (Chapter 11). Their molecular pathogenesis is incompletely understood but seems to involve strain-specific *H. pylori* factors, as well as host genetic factors, such as polymorphisms in the promoters of inflammatory cytokines such as IL-1β and tumor necrosis factor (TNF). It is thought that *H. pylori* infection leads to the activation of *H. pylori*–reactive T cells, which in turn cause polyclonal B cell proliferation. In time, a monoclonal B cell tumor emerges in the proliferating B cells, perhaps as a result of accumulation of mutations in growth regulatory genes. In keeping with this model, early in the course of disease, eradication of *H. pylori* "cures" the lymphoma by removing antigenic stimulus for T cells.

SUMMARY

Helicobacter pylori

- *H. pylori* infection has been implicated in both gastric adenocarcinoma and MALT lymphoma.
- The mechanism of *H. pylori*–induced gastric cancers is multifactorial, including immunologically mediated chronic inflammation, stimulation of gastric cell proliferation, and production of reactive oxygen species that damage DNA. *H. pylori* pathogenicity genes, such as *CagA*, also may contribute by stimulating growth factor pathways.
- It is thought that *H. pylori* infection leads to polyclonal B cell proliferations and that eventually a monoclonal B cell tumor (MALT lymphoma) emerges as a result of accumulation of mutations.

HOST DEFENSE AGAINST TUMORS: TUMOR IMMUNITY

The idea that tumors are not entirely "self" was conceived by Ehrlich, who proposed that immune-mediated recognition of autologous tumor cells may be a "positive mechanism" capable of eliminating transformed cells. Subsequently, Lewis Thomas and Macfarlane Burnet formalized this concept by coining the term *immune surveillance* to refer to recognition and destruction of newly appearing tumor cells, which are seen as foreign by the host immune system. That cancers occur implies that immune surveillance is imperfect; the escape of some tumors from such policing, however, does not preclude the possibility that others may have been aborted. This section addresses certain questions about tumor immunity: What is the nature of tumor antigens? What host effector systems may recognize tumor cells? Is tumor immunity effective against spontaneous neoplasms?

Tumor Antigens

Antigens that elicit an immune response have been demonstrated in many experimentally induced tumors and in some human cancers. Initially, they were broadly classified into two categories based on their patterns of expression: *tumor-specific antigens,* which are present only on tumor cells and not on any normal cells, and *tumor-associated antigens,* which are present on tumor cells and also on some normal cells. This classification, however, is imperfect, because many antigens thought to be tumor-specific turned out to be expressed by some normal cells as well. The modern classification of tumor antigens is based on their molecular structure and source.

An important advance in the field of tumor immunology was the development of techniques for identifying tumor antigens that were recognized by cytotoxic T lymphocytes (CTLs), because CTLs are responsible for the major immune defense mechanism against tumors. As described in Chapter 4, CTLs recognize peptides derived from cytoplasmic proteins that are displayed bound to class I major histocompatibility complex (MHC) molecules.

Described next are the main classes of tumor antigens (Fig. 5–32).

Products of Mutated Oncogenes and Tumor Suppressor Genes

Neoplastic transformation, as discussed, results from genetic alterations, some of which may lead to the expression of cell surface antigens that are seen as non-self by the immune system. Antigens in this category are derived from mutant oncoproteins and tumor suppressor proteins. Unique tumor antigens arise from β-catenin, RAS, p53, and CDK4, for which the encoding genes frequently are mutated in tumors. Because the mutant genes are present only in tumors, their peptides are expressed only in tumor cells. Since many tumors may carry the same mutation, such antigens are shared by different tumors. Although CTLs can be induced against such antigens, they do not appear to elicit protective responses in vivo. In some cases, unmutated oncogenes are overexpressed in tumors. The best

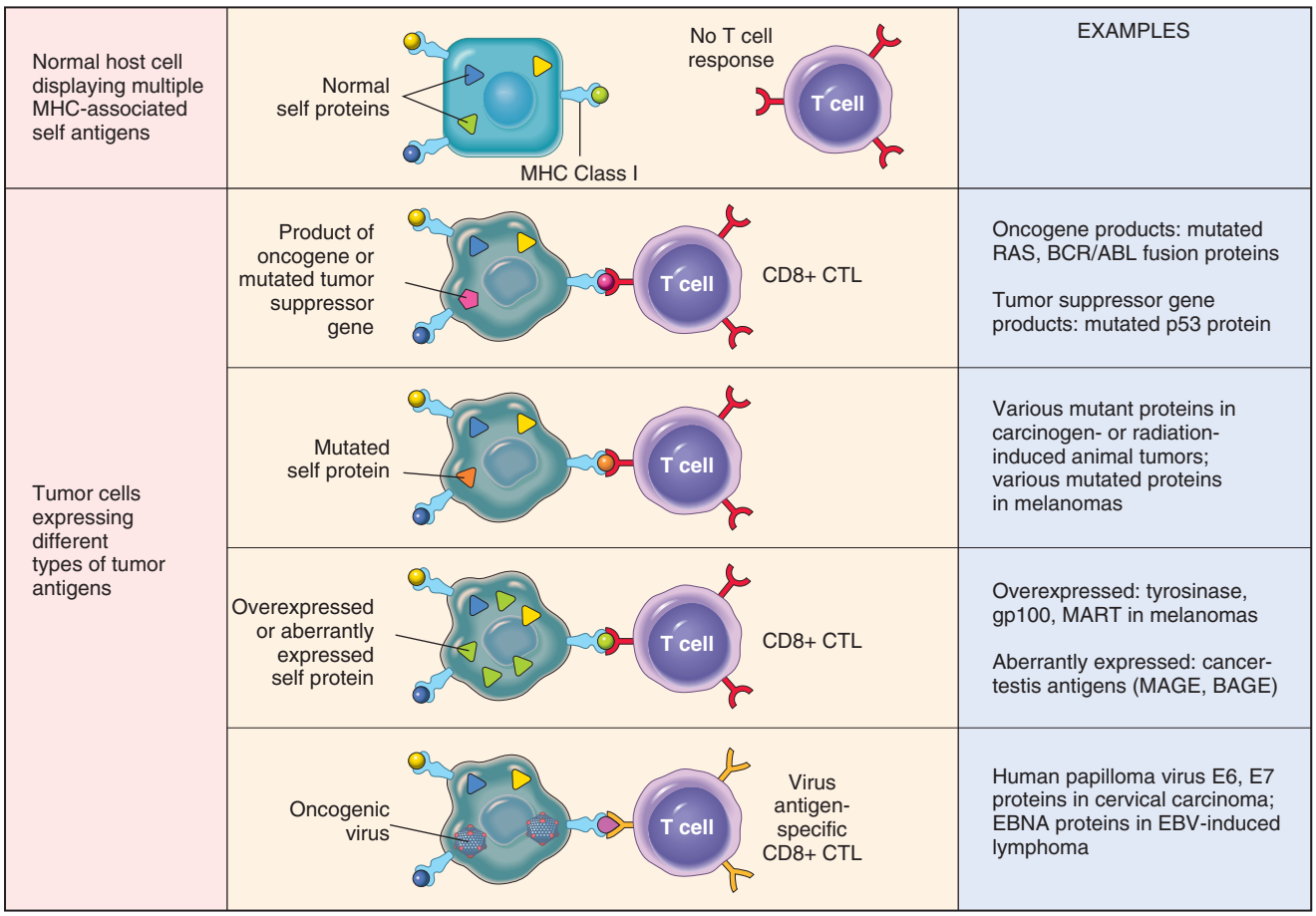


Figure 5–32 Tumor antigens recognized by CD8+ T cells.
(Modified from Abbas AK, Lichtman AH: Cellular and Molecular Immunology, 5th ed. Philadelphia, WB Saunders, 2003.)

example is that of the *HER2/NEU* oncogene, whose product is highly expressed in a subset of breast cancers. Antibodies targeted against Her2/Neu protein are used clinically for the treatment of breast cancers.

Products of Other Mutated Genes

Because of the genetic instability of tumor cells, many genes are mutated in these cells, including genes whose products are not related to the transformed phenotype and have no known function. Products of these mutated genes are potential tumor antigens. These antigens are extremely diverse, because the carcinogens that induce the tumors may randomly mutagenize virtually any host gene. Mutated cellular proteins are found more frequently in chemical carcinogen- or radiation-induced animal tumors than in spontaneous human cancers. They can be targeted by the immune system, since there is no self-tolerance against them.

Overexpressed or Aberrantly Expressed Cellular Proteins

Tumor antigens may be normal cellular proteins that are abnormally expressed in tumor cells and elicit immune responses. In a subset of human melanomas, some tumor antigens are structurally normal proteins that are produced at low levels in normal cells and overexpressed in tumor cells. One such antigen is tyrosinase, an enzyme involved in melanin biosynthesis that is expressed only in normal melanocytes and melanomas. T cells from patients with melanoma recognize peptides derived from tyrosinase, raising the possibility that tyrosinase vaccines may stimulate such responses to melanomas; clinical trials with these vaccines are ongoing. It is somewhat surprising that these patients are able to respond to a normal self-antigen. The probable explanation is that tyrosinase normally is produced in such small amounts and in so few cells that it is not recognized by the immune system and fails to induce tolerance.

Another group, the so-called cancer-testis antigens, are encoded by genes that are silent in all normal adult tissues except the testis, and are deregulated in cancer cells—hence their name. Although the protein is present in the testis, it is not expressed on the cell surface in an antigenic form, because sperm do not express MHC class I molecules. Thus, for all practical purposes, these antigens are tumor-specific. Prototypical of this group is the MAGE (*m*elanoma *a*ntigen *ge*ne) family of genes. Although they are tumor-specific, MAGE antigens are not unique for individual tumors. MAGE-1 is expressed on 37% of melanomas and a variable number of lung, liver, stomach, and esophageal carcinomas. Similar antigens called GAGE, BAGE, and RAGE have been detected in other tumors. Several antigens from this category are now being used in tumor vaccine trials.

Tumor Antigens Produced by Oncogenic Viruses

As discussed earlier, some viruses are associated with cancers. Not surprisingly, these viruses produce proteins that are recognized as foreign by the immune system. The most potent of these antigens are proteins produced by latent DNA viruses; examples in humans are HPV and EBV. There is abundant evidence that CTLs recognize antigens of these viruses and that a competent immune system plays a role in surveillance against virus-induced tumors because of its ability to recognize and kill virus-infected cells. Indeed, vaccines against HPV antigens have been found to be effective in prevention of cervical cancers in girls and young women.

Oncofetal Antigens

Oncofetal antigens or embryonic antigens, such as carcinoembryonic antigen (CEA) and alpha fetoprotein, are expressed during embryogenesis but not in normal adult tissues. Derepression of the genes that encode these antigens causes their reexpression in colon and liver cancers. Antibodies can be raised against these antigens and are useful for detection of oncofetal antigens. Although, as discussed later, they are not entirely tumor-specific, they can serve as serum markers for cancer.

Altered Cell Surface Glycolipids and Glycoproteins

Most human and experimental tumors express higher than normal levels and/or abnormal forms of surface glycoproteins and glycolipids, which may be diagnostic markers and targets for therapy. These altered molecules include gangliosides, blood group antigens, and mucins. Although most of the epitopes recognized by antibodies raised against such antigens are not specifically expressed on tumors, they are present at higher levels on cancer cells than on normal cells. This class of antigens is a target for cancer therapy with specific antibodies.

Several mucins are of special interest and have been the focus of diagnostic and therapeutic studies. These include CA-125 and CA-19-9, expressed on ovarian carcinomas, and MUC-1, expressed on breast carcinomas. Unlike many other types of mucins, MUC-1 is an integral membrane protein that normally is expressed only on the apical surface of breast ductal epithelium, a site that is relatively sequestered from the immune system. In ductal carcinomas of the breast, however, the molecule is expressed in an unpolarized fashion and contains new, tumor-specific carbohydrate and peptide epitopes. These epitopes induce both antibody and T cell responses in cancer patients and are therefore candidates for tumor vaccines.

Cell Type–Specific Differentiation Antigens

Tumors express molecules that normally are present on the cells of origin. These antigens are called *differentiation antigens,* because they are specific for particular lineages or differentiation stages of various cell types. Their importance is as potential targets for immunotherapy and in identifying the tissue of origin of tumors. For example, lymphomas may be diagnosed as B cell–derived tumors by the detection of surface markers characteristic of this lineage, such as CD20. Antibodies against CD20 are used for immunotherapy of certain B cell lymphomas. These differentiation antigens typically are normal self-antigens, so they do not induce immune responses in tumor-bearing hosts.

Antitumor Effector Mechanisms

Cell-mediated immunity is the dominant antitumor mechanism in vivo. Although antibodies can be made against tumors, there is no evidence that they play a protective role under physiologic conditions. The cellular effectors that mediate immunity are discussed fully in Chapter 4, so they are characterized only briefly here.

Cytotoxic T Lymphocytes

The role of specifically sensitized cytotoxic T lymphocytes (CTLs) in experimentally induced tumors is well established. In humans, they seem to play a protective role, chiefly against virus-associated neoplasms (e.g., EBV-induced Burkitt lymphoma, HPV-induced tumors). The presence of MHC-restricted CD8+ cells that can kill autologous tumor cells within human tumors suggests that the role of T cells in immunity against human tumors may be broader than was previously suspected. In some cases, such CD8+ T cells do not develop spontaneously in vivo but can be generated by immunization with tumor antigen–pulsed dendritic cells.

Natural Killer Cells

NK cells are lymphocytes that are capable of destroying tumor cells without previous sensitization; they may provide the first line of defense against tumor cells. After activation with IL-2, NK cells can lyse a wide range of human tumors, including many that seem to be nonimmunogenic for T cells. T cells and NK cells apparently provide complementary antitumor mechanisms. Tumors that fail to express MHC class I antigens cannot be recognized by T cells, but these tumors may trigger NK cells because the latter are inhibited by recognition of normal autologous class I molecules (Chapter 4). Thus, tumors may downregulate MHC class I molecules to avoid recognition by T cells, which then makes them prime targets for NK cells. The triggering receptors on NK cells are extremely diverse and belong to several gene families. NKG2D proteins expressed on NK cells and some T cells are important activating receptors. They recognize stress-induced antigens that are expressed on tumor cells and on cells that have incurred DNA damage and are at risk for neoplastic transformation.

Macrophages

Classically activated macrophages of the M1 type (Chapter 2) exhibit cytotoxicity against tumor cells in vitro. T cells, NK cells, and macrophages may collaborate in antitumor reactivity, because interferon-γ, a cytokine secreted by T cells and NK cells, is a potent activator of macrophages. Activated macrophages may kill tumors by mechanisms similar to those used to kill microbes (e.g., production of reactive oxygen metabolites) (Chapter 2) or by secretion of tumor necrosis factor (TNF).

Humoral Mechanisms

Although there is no evidence for the protective effects of antitumor antibodies against spontaneous tumors,

administration of monoclonal antibodies against tumor cells can be therapeutically effective. A monoclonal antibody against CD20, a B cell surface antigen, is widely used for treatment of certain non-Hodgkin lymphomas.

Immune Surveillance and Immune Evasion by Tumors

In view of the host of possible and potential antitumor mechanisms, is there any evidence that they operate in vivo to prevent the emergence of neoplasms? The strongest argument for the existence of immune surveillance is the increased frequency of cancers in immunodeficient hosts. About 5% of persons with congenital immunodeficiencies develop cancers, a rate that is about 200 times reported rates for persons without such immunodeficiencies. By analogy, immunosuppressed transplant recipients and patients with acquired immunodeficiency syndrome have increased numbers of malignancies. Of note, most (but not all) of these neoplasms are lymphomas, often lymphomas of activated B cells. Particularly illustrative is X-linked lymphoproliferative disorder. When affected boys develop an EBV infection, such infection does not take the usual self-limited form of infectious mononucleosis but instead evolves into a fatal form of infectious mononucleosis or, even worse, malignant lymphoma.

Most cancers occur in persons who do not suffer from any overt immunodeficiency. If immune surveillance exists, how do cancers evade the immune system in immunocompetent hosts? Several escape mechanisms have been proposed:

- *Selective outgrowth of antigen-negative variants.* During tumor progression, strongly immunogenic subclones may be eliminated. This notion is supported by experiments in which tumors arising in immunocompromised mice express antigens that are recognized, with consequent elimination of the tumors by the immune system in normal mice, whereas similar tumors arising in immunocompetent mice are nonimmunogenic.
- *Loss or reduced expression of histocompatibility molecules.* Tumor cells may fail to express normal levels of human leukocyte antigen (HLA) class I, escaping attack by CTLs. Such cells, however, may trigger NK cells.
- *Immunosuppression.* Many oncogenic agents (e.g., chemicals, ionizing radiation) suppress host immune responses. Tumors or tumor products also may be immunosuppressive. For example, TGF-β, secreted in large quantities by many tumors, is a potent immunosuppressant. In some cases, the immune response induced by the tumor may inhibit tumor immunity. Several mechanisms of such inhibition have been described. For instance, recognition of tumor cells may lead to engagement of the T cell inhibitory receptor, CTLA-4, or activation of regulatory T cells that suppress immune responses. More insidiously, some tumors express FasL, which can engage Fas on immune cell surfaces and induce the immune cell to enter apoptosis!
- *Antigen masking*. Many tumor cells produce a thicker coat of external glycocalyx molecules, such as sialic acid–containing mucopolysaccharides, than normal cells. This thick coat may block access of immune cells to antigen-presenting molecules, thereby preventing antigen recognition and cell killing.
- *Downregulation of co-stimulatory molecules.* Costimulatory molecules are required to initiate strong T cell responses. Many tumors reduce expression of these costimulatory molecules.

SUMMARY

Immune Surveillance

- Tumor cells can be recognized by the immune system as non-self and destroyed.
- Antitumor activity is mediated by predominantly cell-mediated mechanisms. Tumor antigens are presented on the cell surface by MHC class I molecules and are recognized by CD8+ CTLs.
- The different classes of tumor antigens include products of mutated proto-oncogenes, tumor suppressor genes, overexpressed or aberrantly expressed proteins, tumor antigens produced by oncogenic viruses, oncofetal antigens, altered glycolipids and glycoproteins, and cell type–specific differentiation antigens.
- Immunosuppressed patients have an increased risk for development of cancer.
- In immunocompetent patients, tumors may avoid the immune system by several mechanisms, including selective outgrowth of antigen-negative variants, loss or reduced expression of histocompatibility antigens, and immunosuppression mediated by secretion of factors (e.g., TGF-β) from the tumor.

CLINICAL ASPECTS OF NEOPLASIA

The importance of neoplasms ultimately lies in their effects on patients. Although malignant tumors are of course more threatening than benign tumors, morbidity and mortality may be associated with any tumor, even a benign one. Indeed, both malignant and benign tumors may cause problems because of (1) location and impingement on adjacent structures, (2) functional activity such as hormone synthesis or the development of paraneoplastic syndromes, (3) bleeding and infections when the tumor ulcerates through adjacent surfaces, (4) symptoms that result from rupture or infarction, and (5) cachexia or wasting. The following discussion considers the effects of a tumor on the host, the grading and clinical staging of cancer, and the laboratory diagnosis of neoplasms.

Effects of Tumor on Host

Location is crucial in both benign and malignant tumors. A small (1-cm) pituitary adenoma can compress and destroy the surrounding normal gland, giving rise to hypopituitarism. A 0.5-cm leiomyoma in the wall of the renal artery may encroach on the blood supply, leading to renal ischemia and hypertension. A comparably small carcinoma within the common bile duct may induce fatal biliary tract obstruction.

Hormone production is seen with benign and malignant neoplasms arising in endocrine glands. Adenomas and carcinomas arising in the beta cells of the pancreatic islets of Langerhans can produce hyperinsulinism, sometimes fatal. By analogy, some adenomas and carcinomas of the adrenal cortex elaborate corticosteroids that affect the patient (e.g., aldosterone, which induces sodium retention, hypertension, and hypokalemia). Such hormonal activity is more likely with a well-differentiated benign tumor than with a corresponding carcinoma.

A tumor may ulcerate through a surface, with consequent bleeding or secondary infection. Benign or malignant neoplasms that protrude into the gut lumen may become caught in the peristaltic pull of the gut, causing intussusception (Chapter 14) and intestinal obstruction or infarction.

Cancer Cachexia

Many cancer patients suffer progressive loss of body fat and lean body mass, accompanied by profound weakness, anorexia, and anemia—a condition referred to as *cachexia.* There is some correlation between the size and extent of spread of the cancer and the severity of the cachexia. However, cachexia is not caused by the nutritional demands of the tumor. Although patients with cancer often are anorexic, current evidence indicates that cachexia results from the action of soluble factors such as cytokines produced by the tumor and the host, rather than reduced food intake. In patients with cancer, calorie expenditure remains high, and basal metabolic rate is increased, despite reduced food intake. This is in contrast with the lower metabolic rate that occurs as an adaptive response in starvation. The basis of these metabolic abnormalities is not fully understood. It is suspected that TNF produced by macrophages in response to tumor cells or by the tumor cells themselves mediates cachexia. TNF suppresses appetite and inhibits the action of lipoprotein lipase, inhibiting the release of free fatty acids from lipoproteins. Additionally, a protein-mobilizing factor called proteolysis-inducing factor, which causes breakdown of skeletal muscle proteins by the ubiquitin-proteosome pathway, has been detected in the serum of cancer patients. Other molecules with lipolytic action also have been found. There is no satisfactory treatment for cancer cachexia other than removal of the underlying cause, the tumor.

Paraneoplastic Syndromes

Symptom complexes that occur in patients with cancer and that cannot be readily explained by local or distant spread of the tumor or by the elaboration of hormones not indigenous to the tissue of origin of the tumor are referred to as *paraneoplastic syndromes.* They appear in 10% to 15% of patients with cancer, and their clinical recognition is important for several reasons:

- Such syndromes may represent the earliest manifestation of an occult neoplasm.
- In affected patients, the pathologic changes may be associated with significant clinical illness and may even be lethal.
- The symptom complex may mimic metastatic disease, thereby confounding treatment.

The paraneoplastic syndromes are diverse and are associated with many different tumors (Table 5–5). *The most common such syndromes are hypercalcemia, Cushing syndrome, and nonbacterial thrombotic endocarditis*; the neoplasms most often associated with these and other syndromes are lung and breast cancers and hematologic malignancies. Hypercalcemia in cancer patients is multifactorial, but the most important mechanism is the synthesis of a parathyroid hormone–related protein (PTHrP) by tumor cells. Also implicated are other tumor-derived factors, such as TGF-α, a polypeptide factor that activates osteoclasts, and the active form of vitamin D. Another possible mechanism for hypercalcemia is widespread osteolytic metastatic disease of bone; of note, however, *hypercalcemia resulting from skeletal metastases is not a paraneoplastic syndrome.* Cushing syndrome arising as a paraneoplastic phenomenon usually is related to ectopic production of ACTH or ACTH-like polypeptides by cancer cells, as occurs in small cell cancers of the lung. Sometimes one tumor induces several syndromes concurrently. For example, bronchogenic carcinomas may elaborate products identical to or having the effects of ACTH, antidiuretic hormone, parathyroid hormone, serotonin, human chorionic gonadotropin, and other bioactive substances.

Paraneoplastic syndromes also may manifest as hypercoagulability, leading to venous thrombosis and nonbacterial thrombotic endocarditis (Chapter 10). Other manifestations are clubbing of the fingers and hypertrophic osteoarthropathy in patients with lung carcinomas (Chapter 12). Still others are discussed in the consideration of cancers of the various organs of the body.

Grading and Staging of Cancer

Methods to quantify the probable clinical aggressiveness of a given neoplasm and its apparent extent and spread in the individual patient are necessary for making an accurate prognosis and for comparing end results of various treatment protocols. For instance, the results of treating extremely small, highly differentiated thyroid adenocarcinomas that are localized to the thyroid gland are likely to be different from those obtained from treating highly anaplastic thyroid cancers that have invaded the neck organs.

The *grading* of a cancer attempts to establish some estimate of its aggressiveness or level of malignancy based on the cytologic differentiation of tumor cells and the number of mitoses within the tumor. The cancer may be classified as grade I, II, III, or IV, in order of increasing anaplasia. Criteria for the individual grades vary with each form of neoplasia and are not detailed here. Difficulties in establishing clear-cut criteria have led in some instances to descriptive characterizations (e.g., "well-differentiated adenocarcinoma with no evidence of vascular or lymphatic invasion" or "highly anaplastic sarcoma with extensive vascular invasion").

Staging of cancers is based on the size of the primary lesion, its extent of spread to regional lymph nodes, and the presence or absence of metastases. This assessment usually is based on clinical and radiographic examination (computed tomography and magnetic resonance imaging) and in some cases surgical exploration. Two methods of

Table 5–5 Paraneoplastic Syndromes

Clinical Syndrome	Major Forms of Neoplasia	Causal Mechanism(s)/Agent(s)
Endocrinopathies		
Cushing syndrome	Small cell carcinoma of lung Pancreatic carcinoma Neural tumors	ACTH or ACTH-like substance
Syndrome of inappropriate antidiuretic hormone secretion	Small cell carcinoma of lung; intracranial neoplasms	Antidiuretic hormone or atrial natriuretic hormones
Hypercalcemia	Squamous cell carcinoma of lung Breast carcinoma Renal carcinoma Adult T cell leukemia/lymphoma Ovarian carcinoma	Parathyroid hormone–related protein, TGF-α, TNF, IL-1
Hypoglycemia	Fibrosarcoma Other mesenchymal sarcomas Hepatocellular carcinoma	Insulin or insulin-like substance
Carcinoid syndrome	Bronchial adenoma (carcinoid) Pancreatic carcinoma Gastric carcinoma	Serotonin, bradykinin
Polycythemia	Renal carcinoma Cerebellar hemangioma Hepatocellular carcinoma	Erythropoietin
Nerve and Muscle Syndrome		
Myasthenia	Bronchogenic carcinoma, thymoma	Immunologic
Disorders of the central and peripheral nervous systems	Breast carcinoma, teratoma	
Dermatologic Disorders		
Acanthosis nigricans	Gastric carcinoma Lung carcinoma Uterine carcinoma	Immunologic; secretion of epidermal growth factor
Dermatomyositis	Bronchogenic and breast carcinoma	Immunologic
Osseous, Articular, and Soft Tissue Changes		
Hypertrophic osteoarthropathy and clubbing of the fingers	Bronchogenic carcinoma	Unknown
Vascular and Hematologic Changes		
Venous thrombosis (Trousseau phenomenon)	Pancreatic carcinoma Bronchogenic carcinoma Other cancers	Tumor products (mucins that activate clotting)
Nonbacterial thrombotic endocarditis	Advanced cancers	Hypercoagulability
Anemia	Thymoma	Immunologic
Others		
Nephrotic syndrome	Various cancers	Tumor antigens, immune complexes

ACTH, adrenocorticotropic hormone; IL-1, interleukin-1; TGF-α, transforming growth factor-α; TNF, tumor necrosis factor.

staging are currently in use: the TNM system (*T,* primary tumor; *N,* regional lymph node involvement; *M,* metastases) and the AJC (American Joint Committee) system. In the *TNM system*, T1, T2, T3, and T4 describe the increasing size of the primary lesion; N0, N1, N2, and N3 indicate progressively advancing node involvement; and M0 and M1 reflect the absence and presence, respectively, of distant metastases. In the *AJC method,* the cancers are divided into stages 0 to IV, incorporating the size of primary lesions and the presence of nodal spread and of distant metastases. Examples of the application of these two staging systems are cited in subsequent chapters. Of note, *when compared with grading, staging has proved to be of greater clinical value.*

SUMMARY

Clinical Aspects of Tumors

- *Cachexia,* defined as progressive loss of body fat and lean body mass, accompanied by profound weakness, anorexia, and anemia, is caused by release of cytokines by the tumor or host.
- Paraneoplastic syndromes, defined as systemic symptoms that cannot be explained by tumor spread or by hormones appropriate to the tissue, are caused by the ectopic production and secretion of bioactive substances such as ACTH, PTHrP, or TGF-α.

- Grading of tumors is determined by cytologic appearance and is based on the idea that behavior and differentiation are related, with poorly differentiated tumors having more aggressive behavior.
- Staging, determined by surgical exploration or imaging, is based on size, local and regional lymph node spread, and distant metastases. Staging is of greater clinical value than grading.

Laboratory Diagnosis of Cancer

Morphologic Methods

In most instances, the laboratory diagnosis of cancer is not difficult. The two ends of the benign–malignant spectrum pose no problems; in the middle, however, lies a "no man's land" where the wise tread cautiously. Clinicians tend to underestimate the contributions they make to the diagnosis of a neoplasm. Clinical and radiologic data are invaluable for optimal pathologic diagnosis. Radiation-induced changes in the skin or mucosa can be similar to those of cancer. Sections taken from a healing fracture can mimic an osteosarcoma. The laboratory evaluation of a lesion can be only as good as the specimen submitted for examination. The specimen must be adequate, representative, and properly preserved.

Several sampling approaches are available, including excision or biopsy, fine-needle aspiration, and cytologic smears. When excision of a lesion is not possible, selection of an appropriate site for biopsy of a large mass requires awareness that the margins may not be representative and the center may be largely necrotic. Requesting *frozen section* diagnosis is sometimes desirable, as, for example, in determining the nature of a mass lesion or in evaluating the regional lymph nodes in a patient with cancer for metastasis. This method, in which a sample is quick-frozen and sectioned, permits histologic evaluation within minutes. In experienced, competent hands, frozen section diagnosis is accurate, but there are particular instances in which the better histologic detail provided by the more time-consuming routine methods is needed. In such instances, it is better to wait a few days, despite the drawbacks, than to perform inadequate or unnecessary surgery.

Fine needle aspiration of tumors is another approach that is widely used. It involves aspiration of cells from a mass, followed by cytologic examination of the smear. This procedure is used most commonly with readily palpable lesions affecting the breast, thyroid, lymph nodes, and salivary glands. Modern imaging techniques permit extension of the method to deeper structures, such as the liver, pancreas, and pelvic lymph nodes. Use of this diagnostic modality obviates surgery and its attendant risks. Although it entails some difficulties, such as small sample size and sampling errors, in experienced hands it can be reliable, rapid, and useful.

Cytologic (Papanicolaou) smears provide another method for the detection of cancer. Historically, this approach has been used widely for discovery of carcinoma of the cervix, often at an in situ stage, but now it is used to investigate many other forms of suspected malignancy, such as endometrial carcinoma, bronchogenic carcinoma, bladder and prostate tumors, and gastric carcinomas; for the identification of tumor cells in abdominal, pleural, joint, and cerebrospinal fluids; and, less commonly, for evaluation of other forms of neoplasia. Neoplastic cells are less cohesive than others and are therefore shed into fluids or secretions (Fig. 5–33). The shed cells are evaluated for features of anaplasia indicative of their origin from a tumor. The gratifying control of cervical cancer is the best testament to the value of the cytologic method.

Immunocytochemistry offers a powerful adjunct to routine histologic examination. Detection of cytokeratin by specific monoclonal antibodies labeled with peroxidase points to a diagnosis of undifferentiated carcinoma rather than large cell lymphoma. Similarly, detection of prostate-specific antigen (PSA) in metastatic deposits by immunohistochemical staining allows definitive diagnosis of a primary tumor in the prostate. Immunocytochemical detection of estrogen receptors allows prognostication and directs therapeutic intervention in breast cancers.

Flow cytometry is used routinely in the classification of leukemias and lymphomas. In this method, fluorescent antibodies against cell surface molecules and

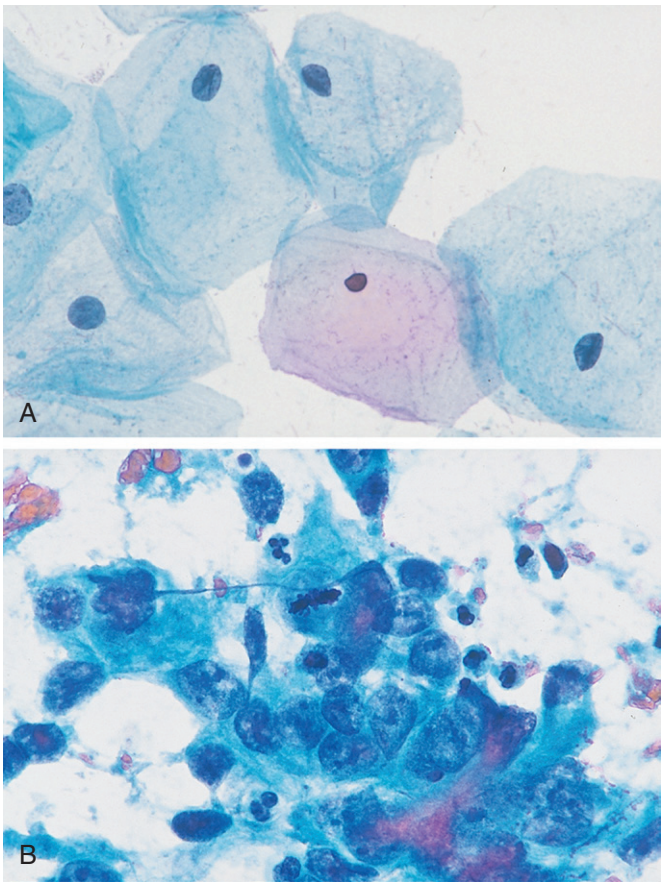


Figure 5–33 A, Normal Papanicolaou smear from the uterine cervix. Large, flat cells with small nuclei are typical. **B,** Abnormal smear containing a sheet of malignant cells with large hyperchromatic nuclei. Nuclear pleomorphism is evident, and one cell is in mitosis. A few interspersed neutrophils, much smaller in size and with compact, lobate nuclei, are seen.

(Courtesy of Dr. Richard M. DeMay, Department of Pathology, University of Chicago, Chicago, Illinois.)

differentiation antigens are used to obtain the phenotype of malignant cells.

Tumor Markers

Biochemical assays for tumor-associated enzymes, hormones, and other tumor markers in the blood cannot be utilized for definitive diagnosis of cancer; however, they can be useful screening tests and in some instances have utility in quantitating the response to therapy or detecting disease recurrence. The application of these assays is considered with many of the specific forms of neoplasia discussed in other chapters, so only a few examples suffice here. PSA, used to screen for prostatic adenocarcinoma, may be one of the most frequently and successfully used tumor markers in clinical practice. Prostatic carcinoma can be suspected when elevated levels of PSA are found in the blood. However, PSA screening also highlights problems encountered with use of virtually every tumor marker. Although PSA levels often are elevated in cancer, PSA levels also may be elevated in benign prostatic hyperplasia (Chapter 17). Furthermore, there is no PSA level that ensures that a patient does not have prostate cancer. *Thus, the PSA test suffers from both low sensitivity and low specificity*. PSA assay is extremely valuable, however, for detecting residual disease or recurrence following treatment for prostate cancer. Other tumor markers occasionally used in clinical practice include carcinoembryonic antigen (CEA), which is elaborated by carcinomas of the colon, pancreas, stomach, and breast, and alpha fetoprotein, which is produced by hepatocellular carcinomas, yolk sac remnants in the gonads, and occasionally teratocarcinomas and embryonal cell carcinomas. Unfortunately, like PSA, both of these markers can be produced in a variety of non-neoplastic conditions as well. Thus, CEA and alpha fetoprotein assays lack both specificity and sensitivity required for the early detection of cancers. As with PSA screening, they are still particularly useful in the detection of recurrences after excision. With successful resection of the tumor, these markers disappear from the serum; their reappearance almost always signifies the beginning of the end. CEA is further discussed in Chapter 14 and alpha fetoprotein in Chapter 15.

Molecular Diagnosis

An increasing number of molecular techniques are being used for the diagnosis of tumors and for predicting their behavior.

- *Diagnosis of malignancy*: Because each T and B cell exhibits unique rearrangement of its antigen receptor genes, polymerase chain reaction (PCR)–based detection of T cell receptor or immunoglobulin genes allows distinction between monoclonal (neoplastic) and polyclonal (reactive) proliferations. Many hematopoietic neoplasms, as well as a few solid tumors, are defined by particular translocations, so the diagnosis can be made by detection of such translocations. For example, fluorescence in situ hybridization (FISH) or PCR analysis (Chapter 6) can be used to detect translocations characteristic of Ewing sarcoma and several leukemias and lymphomas. PCR-based detection of *BCR-ABL* transcripts provides the molecular diagnosis of chronic myeloid leukemia.
- *Prognosis and behavior*: Certain genetic alterations are associated with a poor prognosis, and thus the presence of these alterations determines the patient's subsequent therapy. FISH and PCR methods can be used to detect amplification of oncogenes such as *HER2/NEU* and *NMYC,* which provide prognostic and therapeutic information for breast cancers and neuroblastomas.
- *Detection of minimal residual disease*: Another emerging use of molecular techniques is for detection of minimal residual disease after treatment. For example, detection of *BCR-ABL* transcripts by PCR assay gives a measure of residual disease in patients treated for chronic myeloid leukemia. Recognition that virtually all advanced tumors are associated with both intact circulating tumor cells and products derived from tumors (e.g., tumor DNA) has led to interest in following tumor burden through sensitive blood tests.
- *Diagnosis of hereditary predisposition to cancer*: Germline mutation of several tumor suppressor genes, such as *BRCA1,* increases a patient's risk for development of certain types of cancer. Thus, detection of these mutated alleles may allow the patient and the physician to devise an aggressive screening protocol, as well as an opportunity for prophylactic surgery. In addition, such detection allows genetic counseling of relatives at risk.
- *Therapeutic decision-making*: Therapies that directly target specific mutations are increasingly being developed, and thus detection of such mutations in a tumor can guide the development of targeted therapy, as discussed later. It is now becoming evident that certain targetable mutations may transgress morphologic categories. For example, mutations of the ALK kinase, originally described in a subset of T cell lymphomas, also have been identified in a small percentage of non–small cell carcinomas and neuroblastomas. Clinical trials have shown that lung cancers with ALK mutations respond to ALK inhibitors, whereas other lung cancers do not, leading to recent FDA approval of ALK inhibitors for use in patients with "ALK-mutated" lung cancer. Another recent dramatic example of molecularly "tailored" therapy is seen in melanoma, in which tumors with a valine for glutamate substitution in amino acid 600 (V600E) of the serine/threonine kinase BRAF respond well to BRAF inhibition, whereas melanomas without this mutation show no response. Of some interest, the V600E mutation is also present in a subset of colon cancers, certain thyroid cancers, 100% of hairy cell leukemias, and Langerhans cell histiocytosis (Fig. 5–34). These tumors are morphologically diverse and have distinct cells of origin, but they share identical oncogenic lesions in a common pro-growth pathway.

Molecular Profiling of Tumors

Molecular profiling of tumors can be done both at the level of mRNA and by nucleotide sequencing. Each of these two is described next.

Expression Profiling

This technique allows simultaneous measurements of the expression levels of several thousand genes. The principle of this so-called gene chip technology is illustrated in Figure 5–35 and described briefly here.

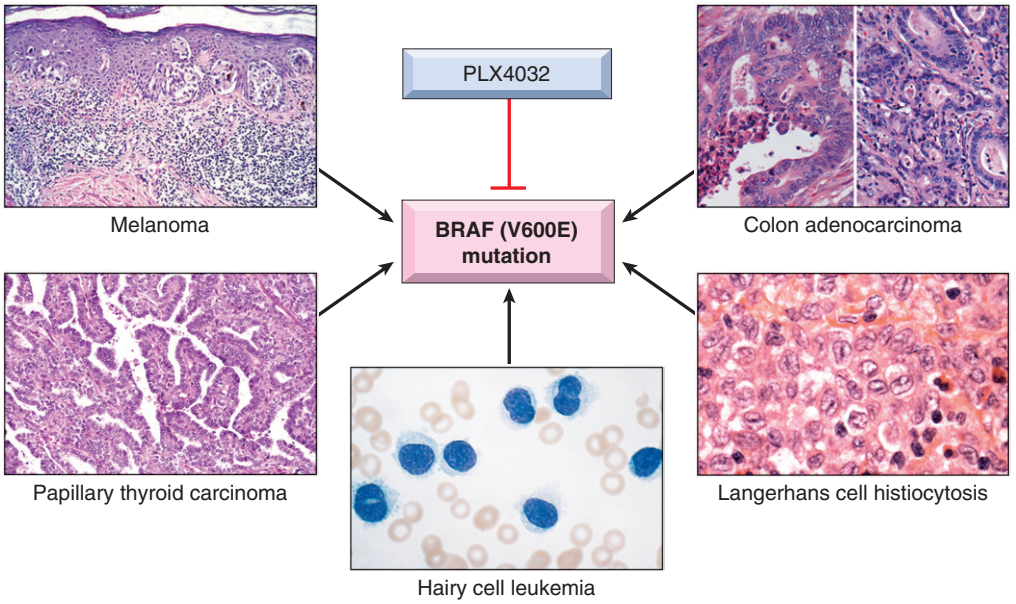


Figure 5–34 Diverse tumor types that share a common mutation, BRAF (V600E), may be candidates for treatments with the same drug, called PLX4032.

As can be seen, the process begins by extraction of mRNA from any two sources (e.g., normal and malignant, normal and preneoplastic, or two tumors of the same histologic type). Complementary DNA (cDNA) copies of the mRNA are synthesized in vitro with fluorescently labeled nucleotides. The fluorescence-labeled cDNA strands are hybridized to sequence-specific DNA probes linked to a solid support, such as a silicon chip. A 1-cm^2 chip can contain thousands of probes arranged in an array of columns and rows. After hybridization, high-resolution laser scanning detects fluorescent signals from each of the spots. The fluorescence intensity of each spot is proportional to the level of expression of the original mRNA used to synthesize the cDNA hybridized to that spot. For each sample, therefore, the expression level of thousands of genes is obtained, and by using bioinformatic tools, the relative levels of gene expression in different samples can be compared. In essence, a molecular profile is generated for each tissue analyzed.

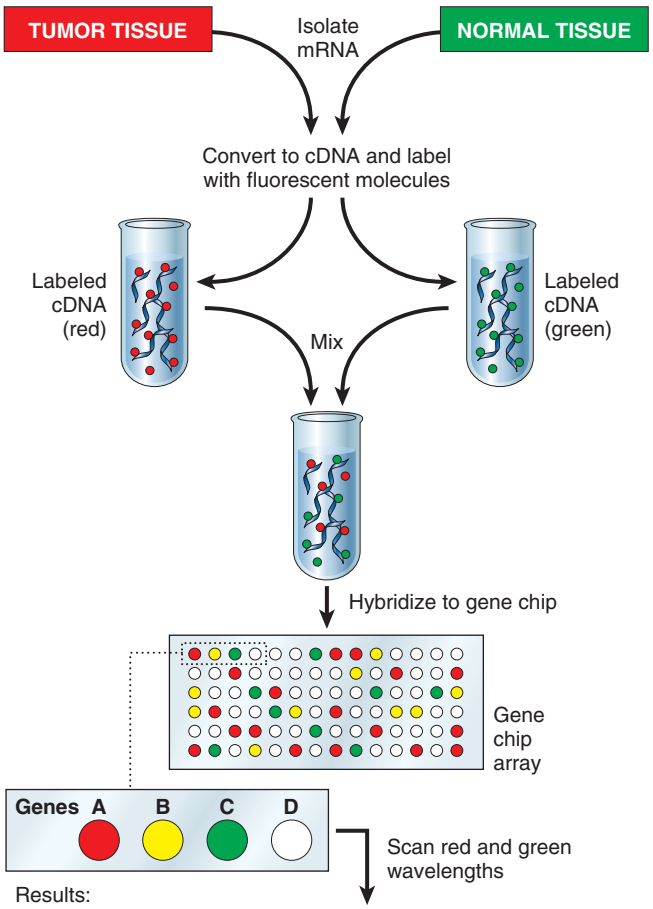
Such analysis has revealed that phenotypically identical large B cell lymphomas (Chapter 11) from different patients are heterogeneous with respect to their gene expression and survival rates. Similar approaches are now being explored in other cancers, such as breast cancers and melanomas.

Whole Genome Sequencing

The progression and development of next-generation sequencing technologies promise even more in-depth analysis of tumors. The advances in such technologies are currently outpacing the famous Moore's law of microprocessors. Sequencing an entire tumor genome, which just a couple of years ago would have taken months and millions of dollars, now takes days and costs a few thousand dollars. Sequences of the entire tumor genomes, when compared with the normal genome from the same patient, can reveal all the somatic alterations present in a tumor.

Recent results from genomic analyses of tumors have revealed that individual tumors can contain from a handful of somatic mutations (certain childhood leukemias) to tens of thousands of mutations, with the highest mutational burden being found in cancers associated with mutagen exposure, such as lung cancer and skin cancer. Among these are two types of mutations: (1) those that subvert normal control of cell proliferation, differentiation, and homeostasis and (2) those that have no effect on cell phenotype. The first set of mutations is referred to as *driver mutations* because they may drive the neoplastic process and hence could be therapeutic targets. The other set of mutations, often much more numerous than driver mutations, most often fall in noncoding regions of the genome or have a neutral effect on growth, not conferring any advantage or disadvantage. Such mutations are called *passenger mutations*. They result from genomic instability of cancer cells and are merely "along for the ride."

In general, driver mutations are recurrent and are present in a substantial percentage of patients with a particular cancer. Thus, for example, *BCR-ABL* fusion genes are present in all cases of chronic myelogenous leukemia, and the fusion protein is an excellent drug target. However, driver mutations may be present in only a subset of tumors of a particular type. For example, approximately 4% of non–small cell lung cancers harbor an *EML4-ALK* tyrosine kinase fusion gene; as already mentioned, in these relatively rare instances, the patient responds well to ALK inhibitors. An additional complication is that some passenger mutations nevertheless have important roles in drug resistance. For example, the mutations in BCR-ABL that confer resistance to imatinib in chronic myelogenous leukemia are present as passenger mutations in rare clones before therapy begins. Because they confer a powerful selective advantage, these mutations are converted from passengers to drivers in the face of drug therapy; it is suspected that the genomic instability of cancer cells sows the seeds of resistance through similar scenarios in many kinds of tumors. Furthermore, in some instances, several distinct and relatively uncommon mutations all converge on the



	Fluorescence intensity Red	Green	Conclusion
Gene A	+++	+	Upregulated in neoplastic tissue
Gene B	++	++	Unchanged in neoplastic tissue
Gene C	+	+++	Downregulated in neoplastic tissue
Gene D	–	–	Not expressed in either tissue

Figure 5–35 Complementary DNA (cDNA) microarray analysis. Messenger RNA (mRNA) is extracted from the samples, reverse transcribed to cDNA, and labeled with fluorescent molecules. In the case illustrated, *red* fluorescent molecules were used for normal cDNA, and *green* molecules were used for tumor cDNA. The labeled cDNAs are mixed and applied to a gene chip, which contains thousands of DNA probes representing known genes. The labeled cDNAs hybridize to spots that contain complementary sequences. The hybridization is detected by laser scanning of the chip, and the results are read in units of red or green fluorescence intensity. In the example shown, *spot A* has high red fluorescence, indicating that a greater number of cDNAs from neoplastic cells hybridized to gene A. Thus, gene A seems to be upregulated in tumor cells. *(Courtesy of Dr. Robert Anders, Department of Pathology, University of Chicago, Chicago, Illinois.)*

same pathway (such as resistance to apoptosis) and contribute to the cancer phenotype. It is therefore useful to categorize mutations on the basis of their ability to drive the cells along the "hallmarks of cancer" pathways.

It is hoped that identification of all potentially targetable mutations in each individual tumor will refocus the treatment of tumors from the tissue of origin to the molecular lesion, as drugs that target specific mutations are developed (Fig. 5–36). This approach represents a paradigm shift

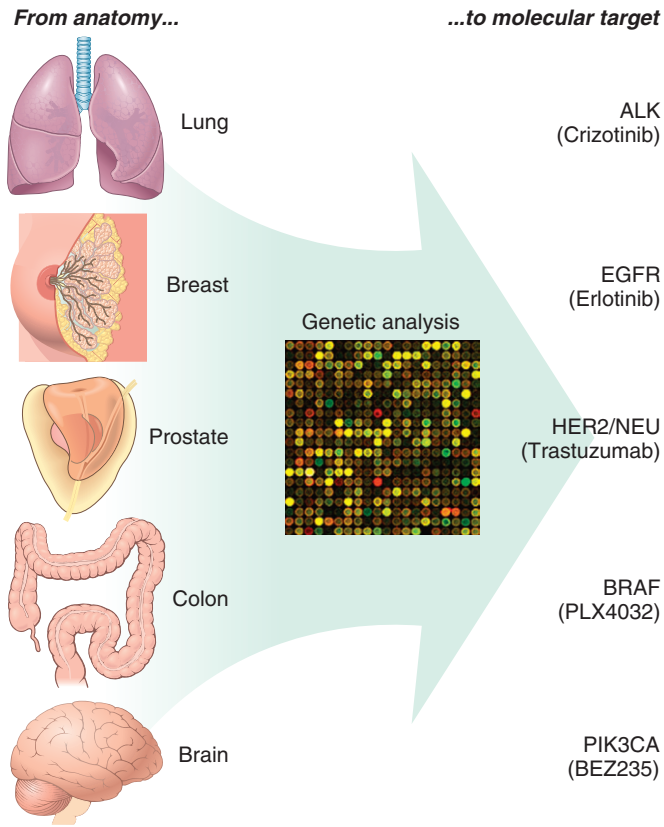


Figure 5–36 A paradigm shift: Classification of cancer according to therapeutic targets rather than cell of origin and morphology.

in the classification and therapy of tumors. Perhaps in the future the diverse group of tumors that bear a common mutation such as BRAF will be classified as BRAF-omas (Fig. 5–34), rather than individual types based on morphology or cell of origin!

SUMMARY

Laboratory Diagnosis of Cancer

- Several sampling approaches exist for the diagnosis of tumors, including excision, biopsy, fine-needle aspiration, and cytologic smears.
- Immunohistochemistry and flow cytometry studies help in the diagnosis and classification of tumors, because distinct protein expression patterns define different entities.
- Proteins released by tumors into the serum, such as PSA, can be used to screen populations for cancer and to monitor for recurrence after treatment.
- Molecular analyses are used to determine diagnosis, prognosis, the detection of minimal residual disease, and the diagnosis of hereditary predisposition to cancer.
- Molecular profiling of tumors by cDNA arrays and sequencing can determine expression of large segments of the genome and catalog all of the mutations in the tumor genome and thus may be useful in molecular stratification of otherwise identical tumors or those of distinct histogenesis that share a mutation for the purpose of treatment and prognostication.

BIBLIOGRAPHY

Ahmed Z, Bicknell R: Angiogenic signalling pathways. Methods Mol Biol 467:3–24, 2009. *[Discussion of many signaling pathways in angiogenesis.]*

Artandi SE, DePinho RA: Telomeres and telomerase in cancer. Carcinogenesis 31:9–18, 2010. *[Review discussing the importance of telomeres and telomerase.]*

Barrallo-Gimeno A, Nieto MA: The Snail genes as inducers of cell movement and survival: implications in development and cancer. Development 132:3151–3161, 2005. *[Discussion of the genes involved in epithelial-mesenchymal transition in cancer.]*

Berx G, van Roy F: Involvement of members of the cadherin superfamily in cancer. Cold Spring Harb Perspect Biol 1:a003129, 2009. *[Review discussing the role of cadherins and contact inhibition in cancer.]*

Bierie B, Moses HL: Tumour microenvironment: TGFbeta: the molecular Jekyll and Hyde of cancer. Nat Rev Cancer 6:506–520, 2006. *[Review discussing the tumor-suppressive and tumor-promoting effects of TGF-β.]*

Burkhart DL, Sage J: Cellular mechanisms of tumour suppression by the retinoblastoma gene. Nat Rev Cancer 8:671–682, 2008. *[Review of* Rb *function.]*

Ciccia A, Elledge SJ: The DNA damage response: making it safe to play with knives. Mol Cell 40:179–204, 2010. *[Review discussing the DNA damage response.]*

Coghlin C, Murray GI: Current and emerging concepts in tumour metastasis. J Pathol 222:1–15, 2010. *[Discussion of current concepts in metastasis.]*

Collado M, Serrano M: Senescence in tumours: evidence from mice and humans. Nat Rev Cancer 10:51–57, 2010. *[Update on mechanisms of senescence.]*

Feron O: Pyruvate into lactate and back: from the Warburg effect to symbiotic energy fuel exchange in cancer cells. Radiother Oncol 92:329–333, 2009. *[An account of the reemergence and molecular pathways of reprogramming of energy metabolism in cancer.]*

Grivennikov SI, Greten FR, Karin M: Immunity, inflammation, and cancer. Cell 140:883–899, 2010. *[A summary of the links between inflammation and the development of cancer.]*

Hanahan D, Weinberg RA: The hallmarks of cancer (2011): the next generation. Cell 144:646–674, 2011. *[Reexamination of the hallmarks of cancer.]*

Junttila MR, Evan GI: p53—a jack of all trades but master of none. Nat Rev Cancer 9:821–829, 2009. *[Update summarizing p53 function.]*

Kalluri R, Zeisberg M: Fibroblasts in cancer. Nat Rev Cancer 6:392–401, 2006. *[Review discussing the role of stroma in cancer.]*

Mathew R, Karantza-Wadsworth V, White E: Role of autophagy in cancer. Nat Rev Cancer 7:961–967, 2007. *[A discussion of the mechanisms of autophagy.]*

Negrini S, Gorgoulis VG, Halazonetis TD: Genomic instability—an evolving hallmark of cancer. Nat Rev Mol Cell Biol 11:220–228, 2010. *[Review on mechanisms of genomic instability, an enabler of malignancy.]*

Perona R: Cell signalling: growth factors and tyrosine kinase receptors. Clin Transl Oncol 8:77–82, 2006. *[Update on signaling pathways in cancer.]*

Stratton MR, Campbell PJ, Futreal PA: The cancer genome. Nature 458:719–724, 2009. *[Excellent summary of next-generation sequencing technologies and their application to cancer.]*

Willis SN, Adams JM: Life in the balance: how BH3-only proteins induce apoptosis. Curr Opin Cell Biol 17:617–625, 2005. *[A review of the mechanisms of apoptosis.]*

Witsch E, Sela M, Yarden Y: Roles for growth factors in cancer progression. Physiology (Bethesda) 25:85–101, 2010. *[An update on the role of growth factors in cancer.]*

See Targeted Therapy available online at studentconsult.com

CHAPTER 6

Genetic and Pediatric Diseases

CHAPTER CONTENTS

GENETIC DISEASES

The completion of the human genome project has been a landmark event in the study of human diseases. It has now been established that humans have only about 25,000 protein-coding genes, far fewer than the 100,000 previously estimated and almost half the number in the lowly rice plant (*Oryza sativa*)! The unraveling of this "genetic architecture" promises to unlock secrets of inherited as well as acquired human disease, since ultimately all diseases involve changes in gene structure or expression. Powerful technologies now allow applications of the human gene sequences to the analysis of human diseases. For example, the human genome project cost approximately 3 billion dollars and many years to complete; current high-throughput sequencing technologies can do the same work in a few weeks for less than $10,000. The speed and reduced costs of DNA sequencing are increasingly facilitating the application of "personalized medicine" to the treatment of cancer and other diseases with a genetic component.

Because several pediatric disorders are of genetic origin, developmental and pediatric diseases are discussed along with genetic diseases in this chapter. However, *it must be borne in mind that not all genetic disorders manifest in infancy and childhood, and conversely, many pediatric diseases are not of genetic origin.* To the latter category belong diseases resulting from immaturity of organ systems. In this context it is helpful to clarify three commonly used terms: hereditary, familial, and congenital. *Hereditary* disorders, by definition, are derived from one's parents, are transmitted in the gametes through the generations, and therefore are *familial.* The term *congenital* simply implies "present at birth." Of note, some congenital diseases are not genetic (e.g., congenital syphilis). On the other hand, not all genetic diseases are congenital; the expression of Huntington

disease, for example, begins only after the third or fourth decade of life.

NATURE OF GENETIC ABNORMALITIES CONTRIBUTING TO HUMAN DISEASE

There are several types of genetic abnormalities that affect the structure and function of proteins, disrupting cellular homeostasis and contributing to disease.

Mutations in Protein-Coding Genes

As is well recognized, the term *mutation* refers to permanent changes in the DNA. Those that affect germ cells are transmitted to the progeny and may give rise to inherited diseases. Mutations in somatic cells are not transmitted to the progeny but are important in the causation of cancers and some congenital malformations.

Details of specific mutations and their effects are discussed along with the relevant disorders throughout this book. Cited here are some common examples of gene mutations and their effects:

- *Point mutations* result from the substitution of a single nucleotide base by a different base, resulting in the replacement of one amino acid by another in the protein product. The mutation in the β-globin chain of hemoglobin giving rise to sickle cell anemia is an excellent example of a point mutation that alters the meaning of the genetic code. Such mutations are sometimes called *missense mutations.*
- By contrast, certain point mutations may change an amino acid codon to a chain termination codon, or *stop codon.* Such "nonsense" mutations interrupt translation, and in most cases RNAs are rapidly degraded, a phenomenon called nonsense mediated decay, such that little or no protein is formed.
- *Frameshift mutations* occur when the insertion or deletion of one or two base pairs alters the reading frame of the DNA strand.
- *Trinucleotide repeat mutations* belong to a special category, because these mutations are characterized by amplification of a sequence of three nucleotides. Although the specific nucleotide sequence that undergoes amplification varies with different disorders, all affected sequences share the nucleotides guanine (G) and cytosine (C). For example, in fragile X syndrome, prototypical of this category of disorders, there are 200 to 4000 tandem repeats of the sequence CGG within a gene called *FMR1.* In normal populations, the number of repeats is small, averaging 29. The expansions of the trinucleotide sequences prevent normal expression of the *FMR1* gene, thus giving rise to mental retardation. Another distinguishing feature of trinucleotide repeat mutations is that they are dynamic (i.e., the degree of amplification increases during gametogenesis). These features, discussed in greater detail later in this chapter, influence the pattern of inheritance and the phenotypic manifestations of the diseases caused by this class of mutations.

Alterations in Protein-Coding Genes Other Than Mutations

In addition to alterations in DNA sequence, coding genes also can undergo structural variations, such as copy number changes (amplifications or deletions), or translocations, resulting in aberrant gain or loss of protein function. As with mutations, structural changes may occur in the germline, or be acquired in somatic tissues. In many instances, pathogenic germ line alterations can involve a contiguous portion of a chromosome rather than a single gene, such as in the 22q microdeletion syndrome, discussed later on. With the widespread availability of array technology for assessing genome-wide DNA copy number variation at very high resolution, pathogenic structural alterations have now been discovered in common disorders such as autism. Cancers often contain somatically acquired structural alterations, including amplifications, deletions, and translocations. The so-called Philadelphia chromosome—translocation t(9;22) between the *BCR* and *ABL* genes in chronic myelogenous leukemia (Chapter 11)—is a classic example.

Sequence and Copy Number Variations (Polymorphisms)

A surprising revelation from the recent progress in genomics is that, on average, any two individuals share greater than 99.5% of their DNA sequences. Thus, the remarkable diversity of humans is encoded in less than 0.5% of our DNA. Though small when compared to the total nucleotide sequences, this 0.5% represents about 15 million base pairs. The two most common forms of DNA variations (polymorphisms) in the human genome are single-nucleotide polymorphisms (SNPs) and copy number variations (CNVs).

- SNPs represent variation at single isolated nucleotide positions and are almost always biallelic (i.e., one of only two choices exist at a given site within the population, such as A or T). Much effort has been devoted to making SNP maps of the human genome. These efforts have identified over 6 million SNPs in the human population, many of which show wide variation in frequency in different populations. SNPs may occur anywhere in the genome—within exons, introns, or intergenic regions—but less than 1% of SNPs occurs in coding regions. These coding sequence variations are important, since they could alter the gene product and predispose to a phenotypic difference or to a disease. Much more commonly, however, the SNP is just a marker that is co-inherited with a disease-associated gene as a result of physical proximity. Another way of expressing this is to say that the SNP and the causative genetic factor are in linkage disequilibrium. There is optimism that groups of SNPs could serve as reliable markers of risk for multigenic complex diseases such as type II diabetes and hypertension, and that by identifying such variants, strategies for disease prevention could be developed (discussed later).
- CNVs are a recently identified form of genetic variation consisting of different numbers of large contiguous stretches of DNA from 1000 base pairs to millions of base pairs. In some instances these loci are, like SNPs, biallelic and simply duplicated or deleted in a subset of

the population. In other instances there are complex rearrangements of genomic material, with multiple alleles in the human population. Current estimates are that CNVs are responsible for between 5 and 24 million base pairs of sequence difference between any two individuals. Approximately 50% of CNVs involve gene-coding sequences; thus, CNVs may underlie a large portion of human phenotypic diversity. There is a significant overrepresentation of certain gene families in regions affected by CNVs; these include genes involved in the immune system and in the nervous system. It is assumed that copy number diversity in such gene families has been subject to strong evolutionary selection, since they would enhance human adaptation to changing environmental factors.

Epigenetic Changes

Epigenetic changes are those involving modulation of gene or protein expression in the absence of alterations in DNA sequence (i.e., mutation) or structure of the encoding gene. Epigenetic regulation is of critical importance during development, as well as in homeostasis of fully developed tissues. One central mechanism of epigenetic regulation is by alterations in the methylation of cytosine residues at gene promoters—heavily methylated promoters become inaccessible to RNA polymerase, leading to transcriptional silencing. Promoter methylation and silencing of tumor suppressor genes (Chapter 5) commonly are observed in many human cancers, leading to unchecked cell growth and proliferation. Another major player in epigenetic regulation of transcription involves the family of *histone proteins,* which are components of structures called nucleosomes, around which DNA is coiled. Histone proteins undergo a variety of reversible modifications (e.g., methylation, acetylation) that affect secondary and tertiary DNA structure, and hence, gene transcription. As expected, abnormalities of histone modification are observed in many acquired diseases such as cancer, leading to transcriptional deregulation. Physiologic epigenetic silencing during development is called *imprinting,* and disorders of imprinting are discussed later on.

Alterations in Non-Coding RNAs

It is worth noting that until recently the major focus of gene hunting has been discovery of genes that encode for proteins. Recent studies indicate, however, that a very large number of genes do not encode proteins. Instead, the non-encoded products of these genes—so-called "non-coding RNAs (ncRNAs)"—play important regulatory functions. Although many distinct families of ncRNAs exist, here we will only discuss two examples: small RNA molecules called *microRNAs (miRNAs),* and *long non-coding RNAs (lncRNAs)* (the latter encompasses ncRNAs >200 nucleotides in length). The miRNAs, unlike messenger RNAs, do not encode proteins but instead inhibit the translation of target mRNAs into their corresponding proteins. Posttranscriptional silencing of gene expression by miRNA is preserved in all living forms from plants to humans and is therefore a fundamental mechanism of gene regulation. Because of their profound influence on gene regulation, miRNAs are assuming central importance in efforts to elucidate normal developmental pathways, as well as pathologic conditions, such as cancer. Andrew Fire and Craig Mello were awarded the Nobel prize in physiology or medicine in 2006 for their work on miRNAs.

By current estimates, there are approximately 1000 genes in humans that encode miRNAs. Transcription of miRNA genes produces primary miRNA transcript (pri-miRNA), which is processed within the nucleus to form another structure called pre-miRNA (Fig. 6–1). With the

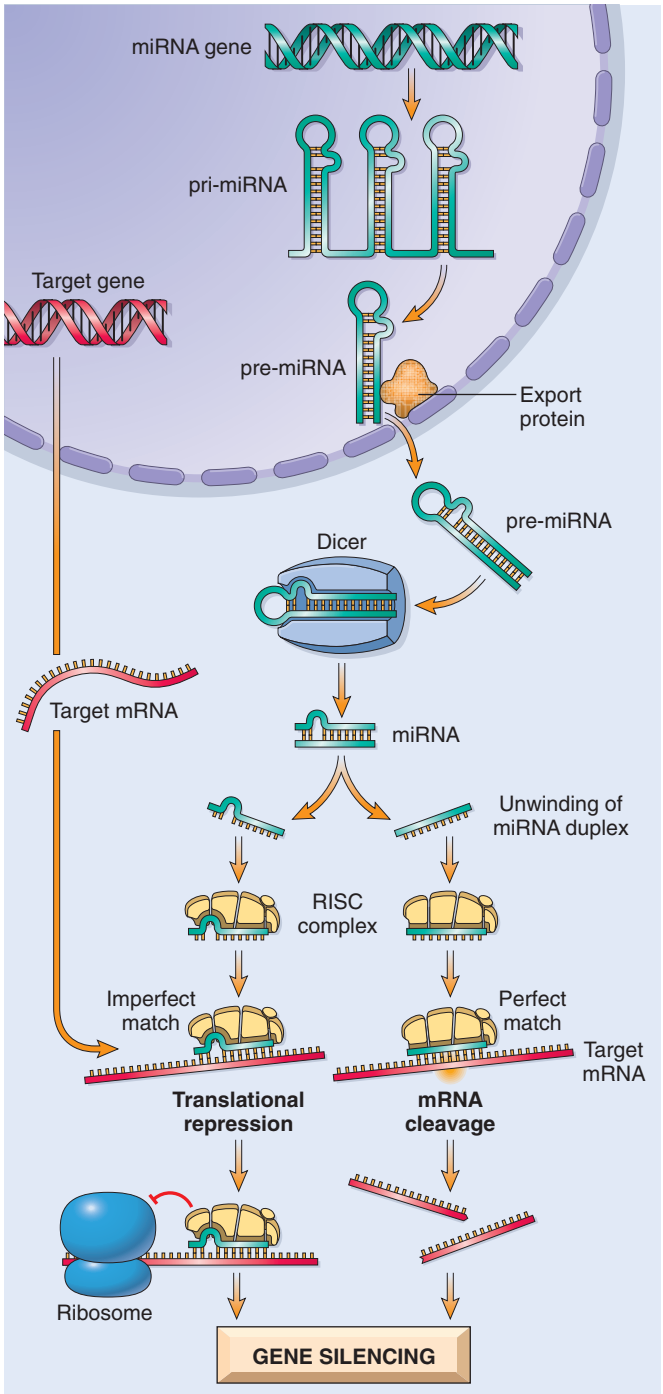


Figure 6–1 Generation of microRNAs and their mode of action in regulating gene function. pri-miRNA, primary microRNA transcript; pre-miRNA, precursor microRNA; RISC, RNA-induced silencing complex.

help of specific transporter proteins, pre-miRNA is exported to the cytoplasm. Additional "cutting" by an enzyme, appropriately called Dicer, generates mature miRNAs that are about 21 to 30 nucleotides in length (hence the designation *micro-*). At this stage the miRNA is still double-stranded. Next, the miRNA unwinds, and single strands of this duplex are incorporated into a multiprotein complex called RNA-induced silencing complex (RISC). Base pairing between the miRNA strand and its target mRNA directs the RISC to either cause mRNA cleavage or repress its translation. In this way, the gene from which the target mRNA was derived is silenced (at a post-transcriptional state). Given that the numbers of miRNA genes are far fewer than those that encode proteins, it follows that a given miRNA can silence many target genes. All mRNAs contain a so-called seed sequence in their 3′ untranslated region (UTR), which determines the specificity of miRNA binding and gene silencing.

Another species of gene-silencing RNA, called small interfering RNAs (siRNAs), works in a manner quite similar to that of miRNA. Unlike miRNA, however, siRNA precursors are introduced by investigators into the cell. Their processing by Dicer and functioning via RISC are essentially similar to that described for miRNA. Synthetic siRNAs have become powerful tools for studying gene function in the laboratory, and are being developed as possible therapeutic agents to silence specific genes, such as oncogenes, whose products are involved in neoplastic transformation.

Recent studies have elucidated an untapped universe of lncRNAs (by some calculations, the number of lncRNAs may exceed coding mRNAs by ten-fold to twenty-fold), and their putative functions in the human genome might explain why humans are at the apex of the evolutionary pyramid despite the relatively modest number of coding genes. lncRNAs modulate gene expression in many ways; for example, they can bind to regions of chromatin, restricting access of RNA polymerase to the encompassed coding genes within the region. One of the best known examples of lncRNAs is XIST, which is transcribed from the X-chromosome, and plays an essential role in physiologic X chromosome inactivation (see later). XIST itself escapes X inactivation, but forms a repressive "cloak" on the X chromosome from which it is transcribed, resulting in gene silencing. Emerging studies are highlighting the roles of lncRNAs in various human diseases, from atherosclerosis to cancer.

With this brief review of the nature of abnormalities that contribute to the pathogenesis of human diseases, we can turn our attention to the three major categories of genetic disorders: (1) those related to mutant genes of large effect, (2) diseases with complex multigenic inheritance (sometimes known as multifactorial disorders), and (3) those arising from chromosomal aberrations. The first category, sometimes referred to as *mendelian disorders,* includes many uncommon conditions, such as the storage diseases and inborn errors of metabolism, all resulting from single-gene mutations of large effect. Most of these conditions are hereditary and familial. The second category includes some of the most common disorders of humans, such as hypertension and diabetes mellitus. Multifactorial, or complex, inheritance implies that both genetic and environmental influences condition the expression of a phenotypic characteristic or disease. The third category includes disorders that are the consequence of numeric or structural abnormalities in the chromosomes.

To these three well-known categories, it is necessary to add a heterogeneous group of genetic disorders that, like mendelian disorders, involve single genes but do not follow simple mendelian rules of inheritance. These single-gene disorders with nonclassic inheritance patterns include those resulting from triplet repeat mutations, those arising from mutations in mitochondrial DNA, and those in which the transmission is influenced by an epigenetic phenomenon called *genomic imprinting*. Each of these four categories is discussed separately.

MENDELIAN DISORDERS: DISEASES CAUSED BY SINGLE-GENE DEFECTS

Single-gene defects (mutations) follow the well-known mendelian patterns of inheritance (Tables 6–1 and 6–2). Although individually each is rare, altogether they account for approximately 1% of all adult admissions to hospitals and about 6% to 8% of all pediatric hospital admissions. Listed next are a few important tenets and caveats of relevance in a consideration of mendelian disorders:

- Mutations involving single genes follow one of three patterns of inheritance: autosomal dominant, autosomal recessive, or X-linked.
- A single-gene mutation may lead to many phenotypic effects (*pleiotropy*), and conversely, mutations at several genetic loci may produce the same trait (*genetic heterogeneity*). For example, Marfan syndrome, which results from a basic defect in connective tissue, is associated

Table 6–1 Estimated Prevalence of Selected Mendelian Disorders Among Liveborn Infants

Disorder	Estimated Prevalence
Autosomal Dominant Inheritance	
Familial hypercholesterolemia	1 in 500
Polycystic kidney disease	1 in 1000
Hereditary spherocytosis	1 in 5000 (Northern Europe)
Marfan syndrome	1 in 5000
Huntington disease	1 in 10,000
Autosomal Recessive Inheritance	
Sickle cell anemia	1 in 500 (U.S. African Americans)*
Cystic fibrosis	1 in 3200 (U.S. Caucasians)
Tay-Sachs disease	1 in 3500 (U.S. Ashkenazi Jewish; French Canadians)
Phenylketonuria	1 in 10,000
Mucopolysaccharidoses—all types	1 in 25,000
Glycogen storage diseases—all types	1 in 50,000
Galactosemia	1 in 60,000
X-Linked Inheritance	
Duchenne muscular dystrophy	1 in 3500 (U.S. males)
Hemophilia	1 in 5000 (U.S. males)

*The prevalence of heterozygous sickle cell *trait* is 1 in 12 for U.S. African Americans.

Table 6–2 Biochemical Basis and Inheritance Pattern for Selected Mendelian Disorders

Disease	Abnormal Protein	Protein Type/Function
Autosomal Dominant Inheritance		
Familial hypercholesterolemia	Low-density lipoprotein receptor	Receptor transport
Marfan syndrome	Fibrillin	Structural support: extracellular matrix
Ehlers-Danlos syndrome*	Collagen	Structural support: extracellular matrix
Hereditary spherocytosis	Spectrin, ankyrin, or protein 4.1	Structural support: red blood cell membrane
Neurofibromatosis, type 1	Neurofibromin-1 (NF-1)	Growth regulation
Adult polycystic kidney disease	Polycystin-1 (PKD-1)	Cell–cell and cell–matrix interactions
Autosomal Recessive Inheritance		
Cystic fibrosis	Cystic fibrosis transmembrane regulator	Ion channel
Phenylketonuria	Phenylalanine hydroxylase	Enzyme
Tay-Sachs disease	Hexosaminidase	Enzyme
Severe combined immunodeficiency	Adenosine deaminase	Enzyme
α- and β-Thalassemias†	Hemoglobin	Oxygen transport
Sickle cell anemia†	Hemoglobin	Oxygen transport
X-linked Recessive Inheritance		
Hemophilia A	Factor VIII	Coagulation
Duchenne/Becker muscular dystrophy	Dystrophin	Structural support: cell membrane
Fragile X syndrome	FMRP	RNA translation

*Some variants of Ehlers-Danlos syndrome have an autosomal recessive inheritance pattern.
†Although full-blown symptoms require biallelic mutations, heterozygotes for thalassemia and sickle cell anemia may present with mild clinical disease. Thus, these disorders sometimes are categorized as "autosomal codominant" entities.

with widespread effects involving the skeleton, eye, and cardiovascular system, all of which stem from a mutation in the gene encoding fibrillin, a component of connective tissues. On the other hand, retinitis pigmentosa, an inherited disorder associated with abnormal retinal pigmentation and consequent visual impairment, can be caused by several different types of mutations. Recognition of genetic heterogeneity not only is important in genetic counseling but also facilitates understanding of the pathogenesis of common disorders such as diabetes mellitus (Chapter 19).

- It is now increasingly being recognized that even known "single-gene" diseases are influenced by inheritance at other genetic loci, which are called *modifier* genes. As discussed later in the section on cystic fibrosis, these modifier loci can affect the severity or extent of the disease.
- The use of proactive prenatal genetic screening in high-risk populations (e.g., persons of Ashkenazi Jewish descent) has significantly reduced the incidence (Table 6–1) of certain genetic disorders such as Tay-Sachs disease.

Transmission Patterns of Single-Gene Disorders

Disorders of Autosomal Dominant Inheritance

Disorders of autosomal dominant inheritance are manifested in the heterozygous state, so at least one parent in an index case usually is affected; both males and females are affected, and both can transmit the condition. When an affected person marries an unaffected one, every child has one chance in two of having the disease. The following features also pertain to autosomal dominant diseases:

- With any autosomal dominant disorder, some patients do not have affected parents. Such patients owe their disorder to new mutations involving either the egg or the sperm from which they were derived. Their siblings are neither affected nor at increased risk for development of the disease.
- Clinical features can be modified by reduced penetrance and variable expressivity. Some persons inherit the mutant gene but are phenotypically normal. This mode of expression is referred to as *reduced penetrance.* The variables that affect penetrance are not clearly understood. In contrast with penetrance, if a trait is consistently associated with a mutant gene but is expressed differently among persons carrying the gene, the phenomenon is called *variable expressivity.* For example, manifestations of neurofibromatosis 1 range from brownish spots on the skin to multiple tumors and skeletal deformities.
- In many conditions, the age at onset is delayed, and symptoms and signs do not appear until adulthood (as in Huntington disease).
- In autosomal dominant disorders, a 50% reduction in the normal gene product is associated with clinical signs and symptoms. Because a 50% loss of enzyme activity can be compensated for, involved genes in autosomal dominant disorders usually do not encode enzyme proteins, but instead fall into two other categories of proteins:
  - Those involved in regulation of complex metabolic pathways, often subject to feedback control (e.g., membrane receptors, transport proteins). An example of this mechanism of inheritance is familial hypercholesterolemia, which results from mutation in the

low-density lipoprotein (LDL) receptor gene (discussed later).
- Key structural proteins, such as collagen and cytoskeletal components of the red cell membrane (e.g., spectrin, abnormalities of which result in hereditary spherocytosis)

The biochemical mechanisms by which a 50% reduction in the levels of such proteins results in an abnormal phenotype are not fully understood. In some cases, especially when the gene encodes one subunit of a multimeric protein, the product of the mutant allele can interfere with the assembly of a functionally normal multimer. For example, the collagen molecule is a trimer in which the three collagen chains are arranged in a helical configuration. Even with a single mutant collagen chain, normal collagen trimers cannot be formed, so there is a marked deficiency of collagen. In this instance the mutant allele is called *dominant negative,* because it impairs the function of a normal allele. This effect is illustrated in some forms of osteogenesis imperfecta (Chapter 20).

Disorders of Autosomal Recessive Inheritance

Disorders of autosomal recessive inheritance make up the largest group of mendelian disorders. They occur when both of the alleles at a given gene locus are mutants; therefore, such disorders are characterized by the following features: (1) The trait does not usually affect the parents, but siblings may show the disease; (2) siblings have one chance in four of being affected (i.e., the recurrence risk is 25% for each birth); and (3) if the mutant gene occurs with a low frequency in the population, there is a strong likelihood that the affected patient (the proband) is the product of a consanguineous marriage.

In contrast with the features of autosomal dominant diseases, the following features generally apply to most autosomal recessive disorders:

- The expression of the defect tends to be more uniform than in autosomal dominant disorders.
- Complete penetrance is common.
- Onset is frequently early in life.
- Although new mutations for recessive disorders do occur, they are rarely detected clinically. Because the affected person is an asymptomatic heterozygote, several generations may pass before the descendants of such a person mate with other heterozygotes and produce affected offspring.
- In many cases, enzyme proteins are affected by the mutation. In heterozygotes, equal amounts of normal and defective enzyme are synthesized. Usually the natural "margin of safety" ensures that cells with half of their complement of the enzyme function normally.

X-Linked Disorders

All sex-linked disorders are X-linked. No Y-linked diseases are known as yet. Save for determinants that dictate male differentiation, the only characteristic that may be located on the Y chromosome is the attribute of hairy ears, which is not altogether devastating. Most X-linked disorders are X-linked *recessive* and are characterized by the following features:

- They are transmitted by heterozygous female carriers only to sons, who of course are hemizygous for the X chromosome.
- Heterozygous females rarely express the full phenotypic change, because they have the paired normal allele; although one of the X chromosomes in females is inactivated (see further on), this process of inactivation is *random,* which typically allows sufficient numbers of cells with the normal expressed allele to emerge.
- An affected male does not transmit the disorder to sons, but all daughters are carriers. Sons of heterozygous women have one chance in two of receiving the mutant gene.

SUMMARY

Transmission Patterns of Single-Gene Disorders

- Autosomal dominant disorders are characterized by expression in heterozygous state; they affect males and females equally, and both sexes can transmit the disorder.
- Enzyme proteins are not affected in autosomal dominant disorders; instead, receptors and structural proteins are involved.
- Autosomal recessive diseases occur when both copies of a gene are mutated; enzyme proteins are frequently involved. Males and females are affected equally.
- X-linked disorders are transmitted by heterozygous females to their sons, who manifest the disease. Female carriers usually are protected because of random inactivation of one X chromosome.

Diseases Caused by Mutations in Genes Encoding Structural Proteins

Marfan Syndrome

In Marfan syndrome, a connective tissue disorder of autosomal dominant inheritance, the basic biochemical abnormality is a mutation affecting fibrillin. This glycoprotein, secreted by fibroblasts, is the major component of microfibrils found in the extracellular matrix. Microfibrils serve as scaffolding for the deposition of tropoelastin, an integral component of elastic fibers. Although microfibrils are widely distributed in the body, they are particularly abundant in the aorta, ligaments, and the ciliary zonules that support the ocular lens; these tissues are prominently affected in Marfan syndrome.

Fibrillin is encoded by the *FBN1* gene, which maps to chromosomal locus 15q21. Mutations in the *FBN1* gene are found in all patients with Marfan syndrome. However, molecular diagnosis of Marfan syndrome is not yet feasible, because more than 600 distinct causative mutations in the very large *FBN1* gene have been found. Since heterozygotes have clinical symptoms, it follows that the mutant fibrillin protein must act as a dominant negative by preventing the assembly of normal microfibrils. The prevalence of Marfan syndrome is estimated to be 1 per 5000. Approximately 70% to 85% of cases are familial, and the rest are sporadic, arising from de novo *FBN1* mutations in the germ cells of parents.

While many of the abnormalities in Marfan syndrome can be explained on the basis of structural failure of connective tissues, some, such as overgrowth of bones, are difficult to relate to simple loss of fibrillin. Recent studies indicate that loss of microfibrils gives rise to abnormal and excessive activation of transforming growth factor-β (TGF-β), since normal microfibrils sequester TGF-β, thereby controlling bioavailability of this cytokine. Excessive TGF-β signaling has deleterious effects on vascular smooth muscle development and the integrity of extracellular matrix. In support of this hypothesis, mutations in the TGF-β type II receptor give rise to a related syndrome, called Marfan syndrome type 2 (MFS2). Of note, angiotensin receptor blockers, which inhibit the activity of TGF-β, have been shown to improve aortic and cardiac function in mouse models of Marfan syndrome and currently are being evaluated in clinical trials.

MORPHOLOGY

Skeletal abnormalities are the most obvious feature of Marfan syndrome. Patients have a slender, elongated habitus with abnormally long legs, arms, and fingers (arachnodactyly); a high-arched palate; and hyperextensibility of joints. A variety of spinal deformities, such as severe kyphoscoliosis, may be present. The chest is deformed, exhibiting either pectus excavatum (i.e., deeply depressed sternum) or a pigeon-breast deformity. The most characteristic **ocular change** is bilateral dislocation, or subluxation, of the lens secondary to weakness of its suspensory ligaments **(ectopia lentis).** This abnormality is so uncommon in persons who do not have this genetic disease that the finding of bilateral ectopia lentis should raise the diagnostic possibility of Marfan syndrome. Most serious, however, is the involvement of the **cardiovascular system.** Fragmentation of the elastic fibers in the tunica media of the aorta predisposes affected patients to aneurysmal dilation and aortic dissection (Chapter 9). These changes, called **cystic medionecrosis,** are not specific for Marfan syndrome. Similar lesions occur in hypertension and with aging. Loss of medial support causes dilation of the aortic valve ring, giving rise to aortic incompetence. The cardiac valves, especially the mitral valve, may be excessively distensible and regurgitant **(floppy valve syndrome),** giving rise to mitral valve prolapse and congestive cardiac failure (Chapter 10). Death from aortic rupture may occur at any age, and aortic rupture is in fact the most common cause of death. Less commonly, cardiac failure is the terminal event.

Although the lesions described are typical of Marfan syndrome, they are not seen in all cases. There is much variation in clinical expression, and some patients may exhibit predominantly cardiovascular lesions with minimal skeletal and ocular changes. The variable expressivity is believed to be related to different allelic mutations in the *FBN1* gene.

Ehlers-Danlos Syndromes

Ehlers-Danlos syndromes (EDSs) are a group of diseases characterized by defects in collagen synthesis or structure. All are single-gene disorders, but the mode of inheritance encompasses both autosomal dominant and recessive patterns. There are approximately 30 distinct types of collagen; all have characteristic tissue distributions and are the products of different genes. To some extent, the clinical heterogeneity of EDS can be explained by mutations in different collagen genes.

At least six clinical and genetic variants of EDS are recognized. Because defective collagen is the basis for these disorders, certain clinical features are common to all variants.

As might be expected, tissues rich in collagen, such as skin, ligaments, and joints, frequently are involved in most variants of EDS. Because the abnormal collagen fibers lack adequate tensile strength, the *skin is hyperextensible and joints are hypermobile.* These features permit grotesque contortions, such as bending the thumb backward to touch the forearm and bending the knee upward to create almost a right angle. Indeed, it is believed that most contortionists have one of the EDSs; however, a predisposition to joint dislocation is one of the prices paid for this virtuosity. *The skin is extraordinarily stretchable, extremely fragile,* and *vulnerable to trauma.* Minor injuries produce gaping defects, and surgical repair or any surgical intervention is accomplished only with great difficulty because of the lack of normal tensile strength. The basic defect in connective tissue may lead to serious internal complications, including rupture of the colon and large arteries (vascular EDS); ocular fragility, with rupture of the cornea and retinal detachment (kyphoscoliotic EDS); and diaphragmatic hernias (classical EDS), among others.

The molecular bases for three of the more common variants are as follows:

- *Deficiency of the enzyme lysyl hydroxylase.* Decreased hydroxylation of lysyl residues in types I and III collagen interferes with the formation of cross-links among collagen molecules. As might be expected, this variant (kyphoscoliotic EDS), resulting from an enzyme deficiency, is inherited as an autosomal recessive disorder.
- *Deficient synthesis of type III collagen resulting from mutations affecting the COL3A1* gene. This variant, the vascular type, is inherited as an autosomal dominant disorder and is characterized by weakness of tissues rich in type III collagen (e.g., blood vessels, bowel wall), predisposing them to rupture.
- *Deficient synthesis of type V collagen* due to mutations in *COL5A1* and *COL5A2* is inherited as an autosomal dominant disorder and results in classical EDS.

SUMMARY

Marfan Syndrome

- Marfan syndrome is caused by a mutation in the *FBN1* gene encoding fibrillin, which is required for structural integrity of connective tissues.
- The major tissues affected are the skeleton, eyes, and cardiovascular system.
- Clinical features may include tall stature, long fingers, bilateral subluxation of lens, mitral valve prolapse, aortic aneurysm, and aortic dissection.
- Clinical trials with drugs that inhibit TGF-β signaling such as angiotensin receptor blockers are ongoing, as these have been shown to improve aortic and cardiac function in mouse models.

Ehlers-Danlos Syndromes

- There are six variants of Ehlers-Danlos syndromes, all characterized by defects in collagen synthesis or assembly. Each of the variants is caused by a distinct mutation.
- Clinical features may include fragile, hyperextensible skin vulnerable to trauma, hypermobile joints, and ruptures involving colon, cornea, or large arteries. Wound healing is poor.

Diseases Caused by Mutations in Genes Encoding Receptor Proteins or Channels

Familial Hypercholesterolemia

Familial hypercholesterolemia is among the most common mendelian disorders; the frequency of the heterozygous condition is 1 in 500 in the general population. It is caused by a mutation in the *LDLR* gene that encodes the receptor for low-density lipoprotein (LDL), the form in which 70% of total plasma cholesterol is transported. A brief review of the synthesis and transport of cholesterol follows.

Normal Cholesterol Metabolism. Cholesterol may be derived from the diet or from endogenous synthesis. Dietary triglycerides and cholesterol are incorporated into chylomicrons in the intestinal mucosa, which drain by way of the gut lymphatics into the blood. These chylomicrons are hydrolyzed by an endothelial lipoprotein lipase in the capillaries of muscle and fat. The chylomicron remnants, rich in cholesterol, are then delivered to the liver. Some of the cholesterol enters the metabolic pool (to be described), and some is excreted as free cholesterol or bile acids into the biliary tract. The endogenous synthesis of cholesterol and LDL begins in the liver (Fig. 6–2). The first step in the synthesis of LDL is the secretion of triglyceride-rich very-low-density lipoprotein (VLDL) by the liver into the blood. In the capillaries of adipose tissue and muscle, the VLDL particle undergoes lipolysis and is converted to intermediate-density lipoprotein (IDL). In comparison with VLDL, the content of triglyceride is reduced and that of cholesteryl esters enriched in intermediate-density lipoprotein (IDL), but IDL retains on its surface two of the three VLDL-associated apolipoproteins B-100 and E. Further metabolism of IDL occurs along two pathways: Most of the IDL particles are directly taken up by the liver through the LDL receptor described later; others are converted to cholesterol-rich LDL by a further loss of triglycerides and apolipoprotein E. In the liver cells, IDL is recycled to generate VLDL.

Two thirds of the resultant LDL particles are metabolized by the LDL receptor pathway, and the rest is metabolized by a receptor for oxidized LDL (scavenger receptor), to be described later. The LDL receptor binds to apolipoproteins B-100 and E and thus is involved in the transport of both LDL and IDL. *Although the LDL receptors are widely distributed, approximately 75% are located on hepatocytes, so the liver plays an extremely important role in LDL metabolism.* The first step in the receptor-mediated transport of LDL involves binding to the cell surface receptor, followed by endocytotic internalization inside so-called "clathrin-coated pits" (Fig. 6–3). Within the cell, the endocytic vesicles fuse with the lysosomes, and the LDL molecule is enzymatically degraded, resulting ultimately in the release of free cholesterol into the cytoplasm. *The cholesterol not only is used by the cell for membrane synthesis but also takes part in intracellular cholesterol homeostasis by a sophisticated system of feedback control*:

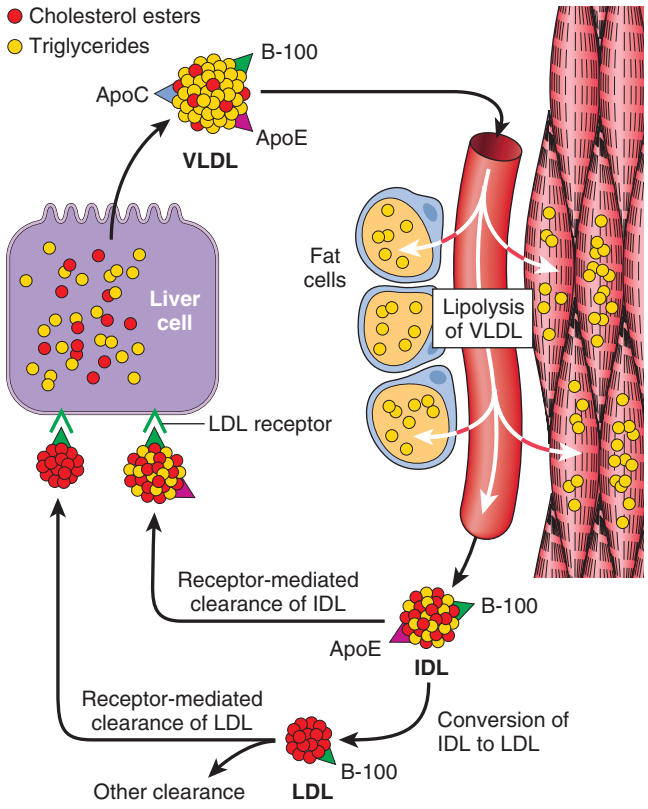


Figure 6–2 Low-density lipoprotein (LDL) metabolism and the role of the liver in its synthesis and clearance. Lipolysis of very-low-density lipoprotein (VLDL) by lipoprotein lipase in the capillaries releases triglycerides, which are then stored in fat cells and used as a source of energy in skeletal muscles. IDL (intermediate-density lipoprotein) remains in the blood and is taken up by the liver.

- It suppresses cholesterol synthesis by inhibiting the activity of the enzyme 3-hydroxy-3-methylglutaryl–coenzyme A reductase (HMG-CoA reductase), which is the rate-limiting enzyme in the synthetic pathway.
- It stimulates the formation of cholesterol esters for storage of excess cholesterol.
- It downregulates the synthesis of cell surface LDL receptors, thus protecting cells from excessive accumulation of cholesterol.

The transport of LDL by the scavenger receptors, alluded to earlier, seems to take place in cells of the mononuclear-phagocyte system and possibly in other cells as well. Monocytes and macrophages have receptors for chemically modified (e.g., acetylated or oxidized) LDLs. The amount catabolized by this "scavenger receptor" pathway is directly related to the plasma cholesterol level.

PATHOGENESIS OF FAMILIAL HYPERCHOLESTEROLEMIA

In familial hypercholesterolemia, mutations in the LDL receptor protein impair the intracellular transport and catabolism

of LDL, resulting in accumulation of LDL cholesterol in the plasma. In addition, the absence of LDL receptors on liver cells also impairs the transport of IDL into the liver, so a greater proportion of plasma IDL is converted into LDL. Thus, patients with familial hypercholesterolemia develop excessive levels of serum cholesterol as a result of the combined effects of reduced catabolism and excessive biosynthesis (Fig. 6–2). In the presence of such hypercholesterolemia, there is a marked increase of cholesterol traffic into the monocyte-macrophages and vascular walls mediated by the scavenger receptor. This accounts for the appearance of skin xanthomas and premature atherosclerosis.

Familial hypercholesterolemia is an autosomal dominant disease. Heterozygotes have a two- to three-fold elevation of plasma cholesterol levels, whereas homozygotes may have in excess of a five-fold elevation. Although their cholesterol levels are elevated from birth, heterozygotes remain asymptomatic until adult life, when they develop cholesterol deposits (xanthomas) along tendon sheaths and premature atherosclerosis resulting in coronary artery disease. Homozygotes are much more severely affected, developing cutaneous xanthomas in childhood and often dying of myocardial infarction before the age of 20 years.

Analysis of the cloned LDL receptor gene has revealed that more than 900 different mutations can give rise to familial hypercholesterolemia. These can be divided into five categories. Class I mutations are uncommon, and they are associated with complete loss of receptor synthesis. With class II mutations, the most prevalent form, the receptor protein is synthesized, but its transport from the endoplasmic reticulum to the Golgi apparatus is impaired due to defects in protein folding. Class III mutations produce receptors that are transported to the cell surface but fail to bind LDL normally. Class IV mutations give rise to receptors that fail to internalize within clathrin pits after binding to LDL, while class V mutations encode receptors that can bind LDL and are internalized but are trapped in endosomes because dissociation of receptor and bound LDL does not occur.

The discovery of the critical role of LDL receptors in cholesterol homeostasis has led to the rational design of the statin family of drugs that are now widely used to lower plasma cholesterol. They inhibit the activity of HMG-CoA reductase and thus promote greater synthesis of LDL receptor (Fig. 6–3).

SUMMARY

Familial Hypercholesterolemia

- Familial hypercholesterolemia is an autosomal dominant disorder caused by mutations in the gene encoding the LDL receptor.
- Patients develop hypercholesterolemia as a consequence of impaired transport of LDL into the cells.
- In heterozygotes, elevated serum cholesterol greatly increases the risk of atherosclerosis and resultant coronary artery disease; homozygotes have an even greater increase in serum cholesterol and a higher frequency of ischemic heart disease. Cholesterol also deposits along tendon sheaths to produce xanthomas.

Cystic Fibrosis

With an incidence of 1 in 3200 live births in the United States, *cystic fibrosis (CF) is the most common lethal genetic disease that affects white populations.* It is uncommon among Asians (1 in 31,000 live births) and African Americans (1 in 15,000 live births). CF follows simple *autosomal recessive* transmission, and does not affect heterozygote carriers. There is, however, a bewildering compendium of phenotypic variation that results from diverse mutations in the CF-associated gene, the tissue-specific effects of loss of this gene's function, and the influence of newly recognized disease modifiers. It is, fundamentally, a *disorder of epithelial transport affecting fluid secretion in exocrine glands and the epithelial lining of the respiratory, gastrointestinal, and reproductive tracts.* Indeed, abnormally viscid mucous secretions that block the airways and the pancreatic ducts are responsible for the two most important clinical manifestations: recurrent and chronic pulmonary infections and pancreatic insufficiency. In addition, although the exocrine sweat glands are structurally normal (and remain so throughout the course of this disease), *a high level of sodium chloride in the sweat is a consistent and characteristic biochemical abnormality in CF.*

PATHOGENESIS

The primary defect in CF is abnormal function of an epithelial chloride channel protein encoded by the CF transmembrane conductance regulator (*CFTR*) gene at chromosomal locus 7q31.2. The changes in mucus are considered secondary to the disturbance in transport of chloride ions. In normal epithelia, the transport of chloride ions across the cell membrane occurs through transmembrane proteins, such as CFTR, that form chloride channels. Mutations in the *CFTR* gene render the epithelial membranes relatively impermeable to chloride ions (Fig. 6–4). However, the impact of this defect on transport function is tissue-specific. The major function of the CFTR protein in the sweat gland ducts is to reabsorb luminal chloride ions and augment sodium reabsorption through the epithelial sodium channel (ENaC). Therefore, in the sweat ducts, loss of CFTR function leads to decreased reabsorption of sodium chloride and production of hypertonic ("salty") sweat (Fig. 6–4, *top*). In contrast with that in the sweat glands, CFTR in the respiratory and intestinal epithelium forms one of the most important avenues for active luminal secretion of chloride. At these sites, *CFTR* mutations result in loss or reduction of chloride secretion into the lumen (Fig. 6–4, *bottom*). Active luminal sodium absorption through ENaCs also is increased, and both of these ion changes increase passive water reabsorption from the lumen, lowering the water content of the surface fluid layer coating mucosal cells. Thus, unlike the sweat ducts, there is no difference in the salt concentration of the surface fluid layer coating the respiratory and intestinal mucosal cells in normal persons and in those with CF. Instead, the pathogenesis of respiratory and intestinal complications in CF seems to stem from an isotonic but low-volume surface fluid layer. In the lungs, this dehydration leads to defective mucociliary action and the accumulation of concentrated, viscid secretions that obstruct the air passages and predispose to recurrent pulmonary infections.

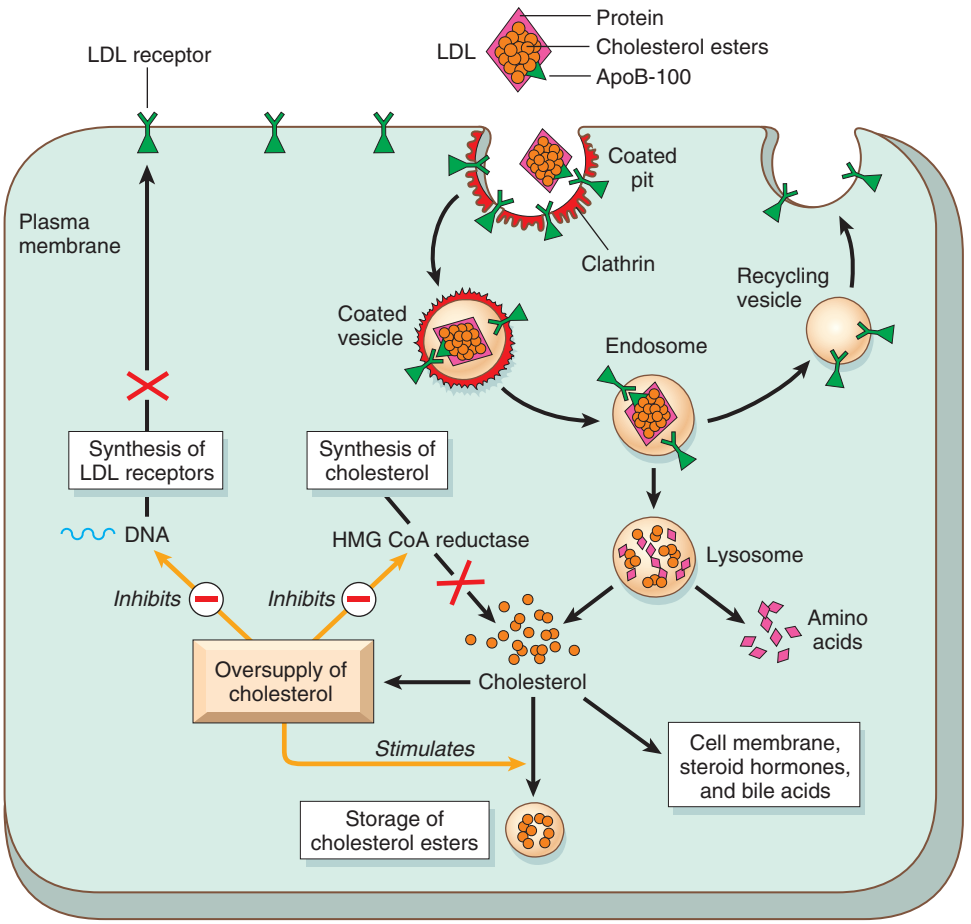


Figure 6–3 The LDL receptor pathway and regulation of cholesterol metabolism. The *yellow arrows* show three regulatory functions of free intracellular cholesterol: (1) suppression of cholesterol synthesis by inhibition of HMG-CoA reductase, (2) stimulating the storage of excess cholesterol as esters, and (3) inhibition of synthesis of LDL receptors. HMG-CoA reductase, 3-hydroxy-3-methylglutaryl–coenzyme A reductase; LDL, low-density lipoprotein.

Since the *CFTR* gene was cloned in 1989, more than 1300 disease-causing mutations have been identified. They can be classified as severe or mild, depending on the clinical phenotype: **Severe** mutations are associated with complete loss of CFTR protein function, whereas **mild** mutations allow some residual function. The most common severe *CFTR* mutation is a deletion of three nucleotides coding for phenylalanine at amino acid position 508 (Δ*F508*). This causes misfolding and total loss of the CFTR. Worldwide, Δ*F508* mutation is found in approximately 70% of patients with CF. Since CF is an autosomal recessive disease, affected persons harbor mutations on both alleles. As discussed later, the combination of mutations on the two alleles influences the overall phenotype, as well as organ-specific manifestations. Although CF remains one of the best-known examples of the "one gene–one disease" axiom, there is increasing evidence that other genes modify the frequency and severity of organ-specific manifestations. One example of a candidate genetic modifier is **mannose-binding lectin,** a key effector of innate immunity involved in phagocytosis of microorganisms. In the setting of CF, polymorphisms in one or both mannose-binding lectin alleles that produce lower circulating levels of the protein are associated with a three-fold higher risk of end-stage lung disease, due to chronic bacterial infections.

MORPHOLOGY

The anatomic changes are highly variable and depend on which glands are affected and on the severity of this involvement. **Pancreatic abnormalities** are present in 85% to 90% of patients with CF. In the milder cases, there may be only accumulations of mucus in the small ducts, with some dilation of the exocrine glands. In more advanced cases, usually seen in older children or adolescents, the ducts are totally plugged, causing atrophy of the exocrine glands and progressive fibrosis (Fig. 6–5). The total loss of pancreatic exocrine secretion impairs fat absorption, so avitaminosis A may contribute to squamous metaplasia of the lining epithelium of the ducts in the pancreas, which are already injured by the inspissated mucus secretions. Thick viscid plugs of mucus also may be found in the small intestine of infants. Sometimes these cause small bowel obstruction, known as **meconium ileus.**

The **pulmonary changes** are the most serious complications of this disease (Fig. 6–6). These changes stem from obstruction and infection of the air passages secondary to the viscous mucus secretions of the submucosal glands of the respiratory tree. The bronchioles often are distended with thick mucus, associated with marked hyperplasia and

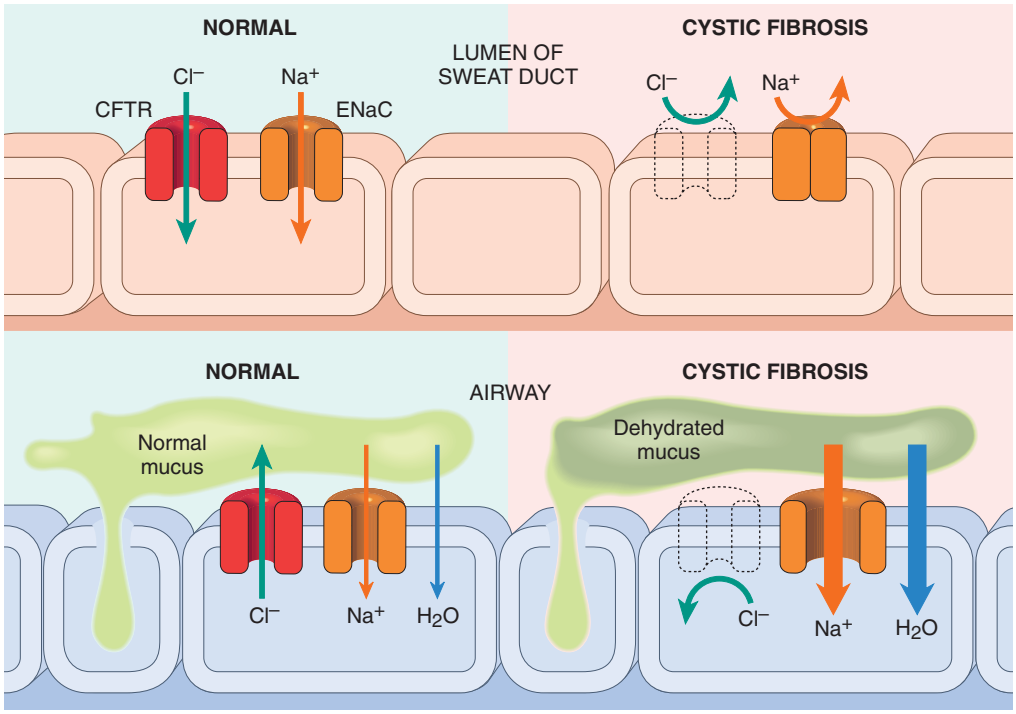


Figure 6–4 Top, In cystic fibrosis (CF), a chloride channel defect in the sweat duct causes increased chloride and sodium concentration in sweat. **Bottom,** Patients with CF have decreased chloride secretion and increased sodium and water reabsorption in the airways, leading to dehydration of the mucus layer coating epithelial cells, defective mucociliary action, and mucous plugging. CFTR, cystic fibrosis transmembrane conductance regulator; ENaC, epithelial sodium channel responsible for intracellular sodium conduction.

hypertrophy of the mucus-secreting cells. Superimposed infections give rise to severe chronic bronchitis and bronchiectasis. Development of lung abscesses is common. *Staphylococcus aureus, Haemophilus influenzae,* and *Pseudomonas aeruginosa* are the three most common organisms responsible for lung infections. Even more sinister is the increasing frequency of infection with another pseudomonad, *Burkholderia cepacia.* This opportunistic bacterium is particularly hardy, and infection with this organism has been associated with fulminant illness ("cepacia syndrome"). The **liver**

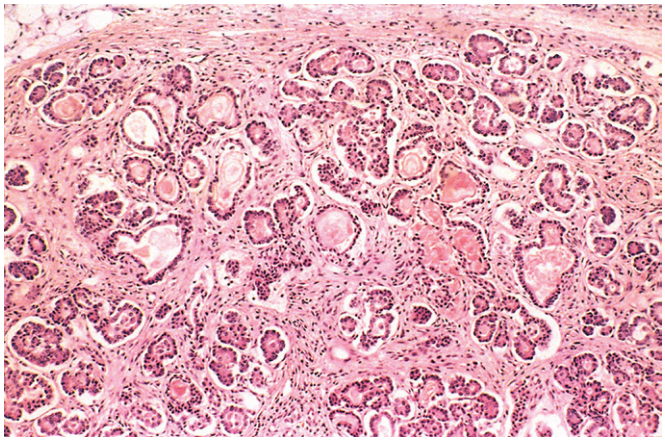


Figure 6–5 Mild to moderate changes of cystic fibrosis in the pancreas. The ducts are dilated and plugged with eosinophilic mucin, and the parenchymal glands are atrophic and replaced by fibrous tissue.

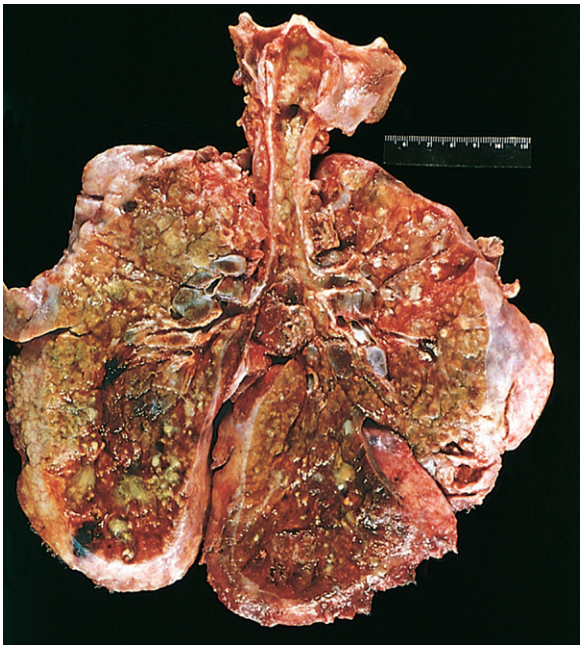


Figure 6–6 Lungs of a patient who died of cystic fibrosis. Extensive mucous plugging and dilation of the tracheobronchial tree are apparent. The pulmonary parenchyma is consolidated by a combination of both secretions and pneumonia; the *greenish* discoloration is the product of *Pseudomonas* infections.

(Courtesy of Dr. Eduardo Yunis, Children's Hospital of Pittsburgh, Pittsburgh, Pennsylvania.)

involvement follows the same basic pattern. Bile canaliculi are plugged by mucinous material, accompanied by ductular proliferation and portal inflammation. Hepatic **steatosis** is a common finding in liver biopsies. Over time, **cirrhosis** develops, resulting in diffuse hepatic nodularity. Such severe hepatic involvement is encountered in less than 10% of patients. **Azoospermia and infertility** are found in 95% of the affected males who survive to adulthood; **bilateral absence of the vas deferens** is a frequent finding in these patients. In some males, this may be the only feature suggesting an underlying *CFTR* mutation.

Clinical Course

In few childhood diseases are clinical manifestations as protean as those of CF (Table 6–3). The signs and symptoms are extremely varied and range from mild to severe, from presence at birth to onset years later, and from confinement to one organ system to involvement of many. Approximately 5% to 10% of the cases come to clinical attention at birth or soon after because of an attack of *meconium ileus. Exocrine pancreatic insufficiency* occurs in a majority (85% to 90%) of patients with CF and is associated with "severe" *CFTR* mutations on *both* alleles (e.g., Δ*F508/*Δ*F508*), whereas 10% to 15% of patients with one "severe" and one "mild" *CFTR* mutation, or two "mild" *CFTR* mutations, retain sufficient pancreatic exocrine function that enzyme supplementation is not required—the *pancreas-sufficient* phenotype. Pancreatic insufficiency is associated with malabsorption of protein and fat and increased fecal loss. Manifestations of malabsorption (e.g., large, foul-smelling stools; abdominal distention; poor weight gain) appear during the first year of life. The faulty fat absorption may induce deficiency states of the fat-soluble vitamins, resulting in manifestations of avitaminosis A, D, or K. Hypoproteinemia may be severe enough to cause generalized edema. Persistent diarrhea may result in rectal prolapse in as many as 10% of children with CF. The pancreas-sufficient phenotype usually is not associated with other gastrointestinal complications, and in general, these patients demonstrate excellent growth and development. *"Idiopathic" chronic pancreatitis* occurs in a subset of patients with pancreas-sufficient CF and is associated with recurring episodes of abdominal pain with life-threatening complications.

Cardiorespiratory complications, such as chronic cough, persistent lung infections, obstructive pulmonary disease, and cor pulmonale, constitute the most common cause of death (accounting for approximately 80% of fatalities) in patients who receive follow-up care in most CF centers in the United States. By 18 years of age, 80% of patients with classic CF harbor *P. aeruginosa,* and 3.5% harbor *B. cepacia.* With the indiscriminate use of antibiotic prophylaxis against *Staphylococcus,* there has been an unfortunate resurgence of resistant strains of *Pseudomonas* in many patients. *Recurrent sinonasal polyps* can occur in as many as 10% to 25% of patients with CF; accordingly, children who present with such polyps should be tested for abnormalities of sweat chloride. Significant *liver disease* occurs late in the natural history of CF and is foreshadowed by pulmonary and pancreatic involvement; with increasing life expectancy, liver disease is now the third most common cause of death in patients with CF (after cardiopulmonary and transplant-related complication).

In most cases, the diagnosis of CF is based on persistently elevated sweat electrolyte concentrations (often the mother makes the diagnosis because her infant "tastes salty"), characteristic clinical findings (sinopulmonary disease and gastrointestinal manifestations), or a family history. Sequencing the *CFTR* gene is, of course, the standard modality for diagnosis of CF. Therefore, in patients with clinical findings or family history (or both) suggestive of this disorder, genetic analysis may be warranted. Advances in management of CF have meant that more patients are now surviving to adulthood; the median life

Table 6–3 Clinical Features and Diagnostic Criteria for Cystic Fibrosis

Clinical Features of Cystic Fibrosis
1. Chronic sinopulmonary disease manifested by
a. Persistent colonization/infection with typical cystic fibrosis pathogens, including *Staphylococcus aureus,* nontypable *Haemophilus influenzae,* mucoid and nonmucoid *Pseudomonas aeruginosa, Burkholderia cepacia*
b. Chronic cough and sputum production
c. Persistent chest radiograph abnormalities (e.g., bronchiectasis, atelectasis, infiltrates, hyperinflation)
d. Airway obstruction manifested by wheezing and air trapping
e. Nasal polyps; radiographic or computed tomographic abnormalities of paranasal sinuses
f. Digital clubbing
2. Gastrointestinal and nutritional abnormalities, including
a. Intestinal: meconium ileus, distal intestinal obstruction syndrome, rectal prolapse
b. Pancreatic: pancreatic insufficiency, recurrent acute pancreatitis, chronic pancreatitis
c. Hepatic: chronic hepatic disease manifested by clinical or histologic evidence of focal biliary cirrhosis, or multilobular cirrhosis, prolonged neonatal jaundice
d. Nutritional: failure to thrive (protein–calorie malnutrition), hypoproteinemia, edema, complications secondary to fat-soluble vitamin deficiency
3. Salt loss syndromes: acute salt depletion, chronic metabolic alkalosis
4. Male urogenital abnormalities resulting in obstructive azoospermia (congenital bilateral absence of vas deferens)
Criteria for Diagnosis of Cystic Fibrosis
One or more characteristic phenotypic features, OR a history of cystic fibrosis in a sibling, OR a positive newborn screening test result
AND
An increased sweat chloride concentration on two or more occasions, OR identification of two cystic fibrosis mutations, OR demonstration of abnormal epithelial nasal ion transport

Adapted with permission from Rosenstein BJ, Cutting GR: The diagnosis of cystic fibrosis: a consensus statement. J Pediatr 132:589, 1998.

expectancy is now 36 years and continues to increase. Clinical trials with gene therapy in humans are still in their early stages but provide a source of encouragement for millions of patients with CF worldwide.

SUMMARY

Cystic Fibrosis

- CF is an autosomal recessive disease caused by mutations in the *CFTR* gene encoding the CF transmembrane regulator.
- The principal defect is of chloride ion transport, resulting in high salt concentrations in sweat and in viscous luminal secretions in respiratory and gastrointestinal tracts.
- *CFTR* mutations can be severe (Δ*F508*), resulting in multisystem disease, or mild, with limited disease extent and severity.
- Cardiopulmonary complications constitute the most common cause of death; pulmonary infections, especially with resistant pseudomonads, are frequent. Bronchiectasis and right-sided heart failure are long-term sequelae.
- Pancreatic insufficiency is extremely common; infertility caused by congenital bilateral absence of vas deferens is a characteristic finding in adult patients with CF.
- Liver disease, including cirrhosis, is increasing in frequency due to improved survival.

Diseases Caused by Mutations in Genes Encoding Enzyme Proteins

Phenylketonuria

There are several variants of phenylketonuria (PKU), an inborn error of metabolism that affects 1 in 10,000 live-born white infants. The most common form, referred to as *classic phenylketonuria*, is quite common in persons of Scandinavian descent and is distinctly uncommon in African American and Jewish populations.

Homozygotes with this autosomal recessive disorder classically have a severe lack of the enzyme phenylalanine hydroxylase (PAH), leading to hyperphenylalaninemia and PKU. Affected infants are normal at birth but within a few weeks exhibit a rising plasma phenylalanine level, which in some way impairs brain development. Usually, by 6 months of life, *severe mental retardation* becomes all too evident; less than 4% of untreated phenylketonuric children have intelligence quotients (IQs) greater than 50 or 60. About one third of these children are never able to walk, and two thirds cannot talk. *Seizures,* other neurologic abnormalities, *decreased pigmentation of hair and skin,* and *eczema* often accompany the *mental retardation* in untreated children. Hyperphenylalaninemia and the resultant mental retardation can be avoided by restriction of phenylalanine intake early in life. Hence, several screening procedures are routinely performed to detect PKU in the immediate postnatal period.

Many female patients with PKU who receive dietary treatment beginning early in life reach child-bearing age and are clinically normal. Most of them have marked hyperphenylalaninemia, because dietary treatment is discontinued after they reach adulthood. Between 75% and 90% of children born to such women are mentally retarded and microcephalic, and 15% have congenital heart disease, even though the infants themselves are heterozygotes. This syndrome, termed *maternal PKU,* results from the teratogenic effects of phenylalanine or its metabolites that cross the placenta and affect specific fetal organs during development. The presence and severity of the fetal anomalies directly correlate with the maternal phenylalanine level, so *it is imperative that maternal dietary restriction of phenylalanine be initiated before conception and continued throughout pregnancy.*

The biochemical abnormality in PKU is an inability to convert phenylalanine into tyrosine. In normal children, less than 50% of the dietary intake of phenylalanine is necessary for protein synthesis. The remainder is converted to tyrosine by the phenylalanine hydroxylase system (Fig. 6–7). When phenylalanine metabolism is blocked because of a lack of PAH enzyme, minor shunt pathways come into play, yielding several intermediates that are excreted in large amounts in the urine and in the sweat. These impart a *strong musty or mousy odor* to affected infants. It is believed that excess phenylalanine or its metabolites contribute to the brain damage in PKU. Concomitant lack of tyrosine (Fig. 6–7), a precursor of melanin, is responsible for the light color of hair and skin.

At the molecular level, approximately 500 mutant alleles of the *PAH* gene have been identified, only some of which cause a severe deficiency of the enzyme. Infants with mutations resulting in a lack of PAH activity present with the classic features of PKU, while those with approximately 6% residual activity present with milder disease. Moreover,

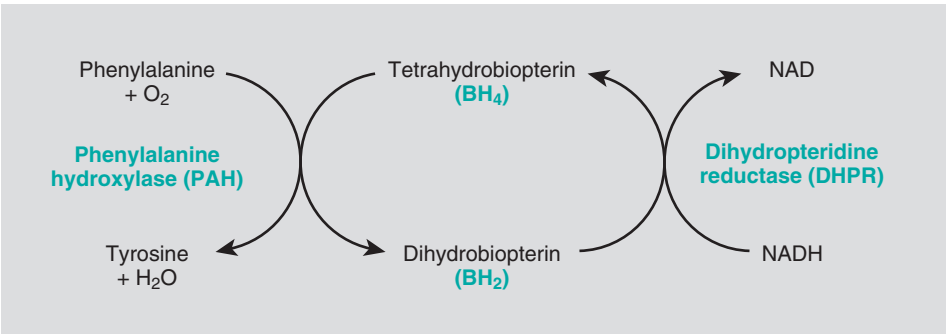


Figure 6–7 The phenylalanine hydroxylase system. NADH, nicotinamide adenine dinucleotide, reduced form.

some mutations result in only modest elevations of blood phenylalanine levels without associated neurologic damage. This latter condition, referred to as *benign hyperphenylalaninemia,* is important to recognize, because affected persons may well have positive screening tests but do not acquire the stigmata of PKU. Because of the numerous disease-causing alleles of the phenylalanine hydroxylase gene, molecular diagnosis is not feasible, and measurement of serum phenylalanine levels is necessary to differentiate benign hyperphenylalaninemia from PKU; the levels in the latter disorder typically are five times (or more) higher than normal. Once a biochemical diagnosis is established, the specific mutation causing PKU can be determined. With this information, carrier testing of at-risk family members can be performed.

While 98% of cases of PKU are attributable to mutations in PAH, approximately 2% arise from abnormalities in synthesis or recycling of the cofactor *tetrahydrobiopterin* (Fig. 6–7). *Clinical recognition of these variant forms of PKU is important to establish a prognosis, because the patients cannot be treated by dietary restriction of phenylalanine.*

Galactosemia

Galactosemia is an autosomal recessive disorder of galactose metabolism that affects 1 in 60,000 live-born infants. Normally, lactase splits lactose, the major carbohydrate of mammalian milk, into glucose and galactose in the intestinal microvilli. Galactose is then converted to glucose in several steps, in one of which the enzyme galactose-1-phosphate uridyltransferase (GALT) is required. Lack of this enzyme, due to homozygous mutations in the encoding gene *GALT*, is responsible for galactosemia. As a result of this transferase deficiency, galactose-1-phosphate and other metabolites, including galactitol, accumulate in many tissues, including the liver, spleen, lens of the eye, kidney, and cerebral cortex.

The liver, eyes, and brain bear the brunt of the damage. The early-onset hepatomegaly is due largely to fatty change, but in time widespread scarring that closely resembles the cirrhosis of alcohol abuse may supervene (Chapter 15). Opacification of the lens (cataract) develops, probably because the lens absorbs water and swells as galactitol, produced by alternative metabolic pathways, accumulates and increases its tonicity. Nonspecific alterations appear in the central nervous system (CNS), including loss of nerve cells, gliosis, and edema. There is still no clear understanding of the mechanism of injury to the liver and brain.

Almost from birth, affected infants fail to thrive. *Vomiting and diarrhea* appear within a few days of milk ingestion. *Jaundice* and *hepatomegaly* usually become evident during the first week of life. Accumulation of galactose and galactose-1-phosphate in the kidney impairs amino acid transport, resulting in aminoaciduria. Fulminant *Escherichia coli* septicemia occurs with increased frequency. The diagnosis of galactosemia can be suspected from demonstration in the urine of a reducing sugar other than glucose, but tests that directly identify the deficiency of the transferase in leukocytes and red cells are more reliable. Antenatal diagnosis is possible by assay of GALT activity in cultured amniotic fluid cells or determination of galactitol level in amniotic fluid supernatant.

Many of the clinical and morphologic changes of galactosemia can be prevented or ameliorated by early removal of galactose from the diet for at least the first 2 years of life. Control instituted soon after birth prevents the cataracts and liver damage and permits almost normal development. Even with dietary restrictions, however, it is now established that older patients frequently are affected by a speech disorder and gonadal failure (especially premature ovarian failure) and, less commonly, by an ataxic condition.

SUMMARY

Phenylketonuria

- PKU is a disorder of autosomal recessive inheritance caused by a lack of the enzyme phenylalanine hydroxylase and consequent inability to metabolize phenylalanine.
- Clinical features of untreated PKU may include severe mental retardation, seizures, and decreased pigmentation of skin, which can be avoided by restricting the intake of phenylalanine in the diet.
- Female patients with PKU who discontinue dietary treatment can give birth to children with malformations and neurologic impairment resulting from transplacental passage of phenylalanine metabolites.

Galactosemia

- Galactosemia is caused by an inherited lack of the GALT enzyme, leading to accumulation of galactose-1-phosphate and its metabolites in tissues.
- Clinical features may include jaundice, liver damage, cataracts, neural damage, vomiting and diarrhea, and *E. coli* sepsis. Dietary restriction of galactose can prevent at least some of the more severe complications.

Lysosomal Storage Diseases

Lysosomes, the digestive system of the cells, contain a variety of hydrolytic enzymes that are involved in the breakdown of complex substrates, such as sphingolipids and mucopolysaccharides, into soluble end products. These large molecules may be derived from the turnover of intracellular organelles that enter the lysosomes by autophagy, or they may be acquired from outside the cell by phagocytosis. With an inherited lack of a lysosomal enzyme, catabolism of its substrate remains incomplete, leading to accumulation of the partially degraded insoluble metabolites within the lysosomes (Fig. 6–8). Approximately 40 lysosomal storage diseases have been identified, each resulting from the functional absence of a specific lysosomal enzyme or proteins involved in their function. Traditionally, lysosomal storage disorders are divided into broad categories based on the biochemical nature of the substrates and the accumulated metabolites, but a more mechanistic classification is based on the underlying molecular defect (Table 6–4). Within each group are several entities, each resulting from the deficiency of a specific enzyme. Despite this complexity, certain features are common to most diseases in this group:

- Autosomal recessive transmission
- Patient population consisting of infants and young children

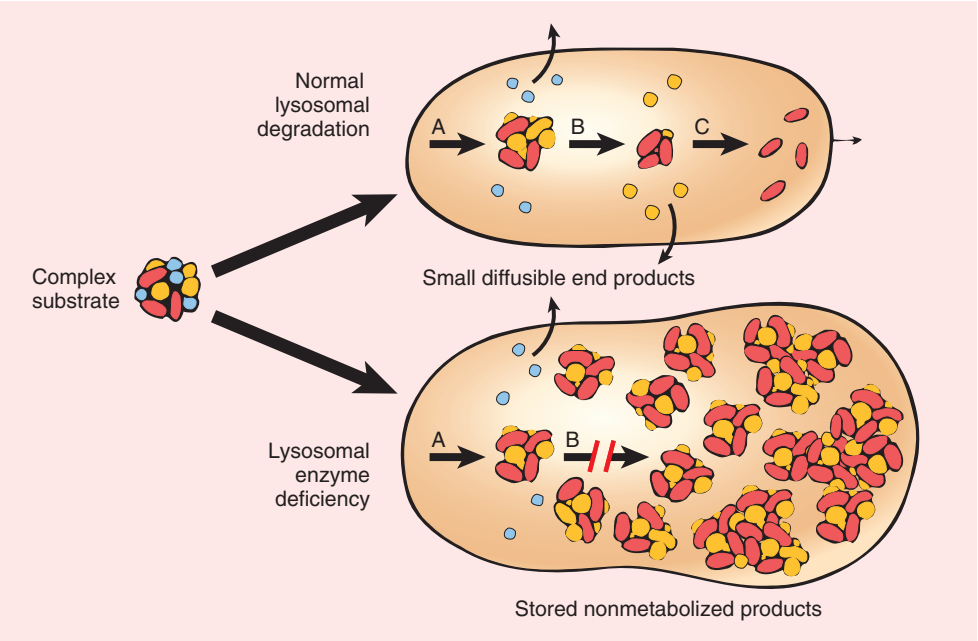


Figure 6–8 Pathogenesis of lysosomal storage diseases. In this example, a complex substrate is normally degraded by a series of lysosomal enzymes (*A*, *B*, and *C*) into soluble end products. If there is a deficiency or malfunction of one of the enzymes (e.g., *B*), catabolism is incomplete, and insoluble intermediates accumulate in the lysosomes.

- Storage of insoluble intermediates in the mononuclear phagocyte system, giving rise to hepatosplenomegaly
- Frequent CNS involvement with associated neuronal damage
- Cellular dysfunctions, caused not only by storage of undigested material but also by a cascade of secondary events triggered, for example, by macrophage activation and release of cytokines

Fortunately for the potential victims of the diseases, most of these conditions are very rare, and their detailed description is better relegated to specialized texts and reviews. Only a few of the more common conditions are considered here. Type II glycogen storage disease (Pompe disease), also a lysosomal disorder, is discussed later in the chapter.

Tay-Sachs Disease (G_{M2} Gangliosidosis: Deficiency in Hexosaminidase β Subunit)

Gangliosidoses are characterized by accumulation of gangliosides, principally in the brain, as a result of a deficiency of a catabolic lysosomal enzyme. Depending on the ganglioside involved, these disorders are subclassified into G_{M1} and G_{M2} categories. Tay-Sachs disease, by far the most common of all gangliosidoses, is characterized by a mutation in and consequent deficiency of the β subunit of the

Table 6–4 Lysosomal Storage Disorders

Disease Category	Disease	Deficiency
Primary lysosomal hydrolase defect	Gaucher disease	Glucocerebrosidase
	G_{M1} gangliosidosis	G_{M1}-β-galactosidase
	Tay-Sachs disease	Hexosaminidase, α subunit
	Sandhoff disease	Hexosaminidase, β subunit
	Fabry disease	α-Galactosidase A
	Krabbe disease	Galactosylceramidase
	Niemann-Pick disease types A and B	Sphingomyelinase
Posttranslational processing defect of lysosomal enzymes	Mucosulfatidosis (juvenile sulfatidosis)	Multiple sulfatases
Inefficient targeting of synthesized hydrolase to the lysosome	Mucolipidosis types II and III alpha/beta	*N*-acetyl glucosamine-1-phosphotransferase
Defect in lysosomal enzyme protection	Galactosialidosis	Protective protein cathepsin A (β-galactosidase and neuraminidase)
Defect in soluble nonenzymatic lysosomal proteins	G_{M2} activator protein deficiency, variant AB	G_{M2} activator protein
	Sphingolipid activator protein deficiency	Sphingolipid activator protein
Transmembrane (nonenzymatic) protein deficiency	Niemann-Pick disease type C (NPC)	*NPC1* and *NPC2*
	Salla disease (free sialic acid storage)	Sialin

Data from Jeyakumar M, Dwek RA, Butters TD, Platt FM: Storage solutions: treating lysosomal disorders of the brain. Nat Rev Neurosci 6:1, 2005.

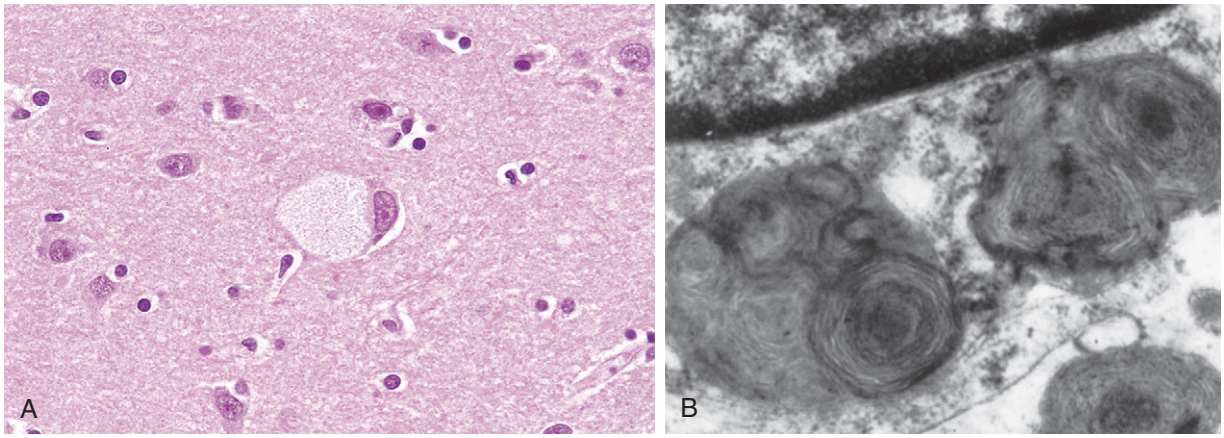


Figure 6–9 Ganglion cells in Tay-Sachs disease. **A,** Under the light microscope, a large neuron has obvious lipid vacuolation. **B,** A portion of a neuron under the electron microscope shows prominent lysosomes with whorled configurations. Part of the nucleus is shown above.
(A, Courtesy of Dr. Arthur Weinberg, Department of Pathology, University of Texas Southwestern Medical Center, Dallas, Texas. B, Courtesy of Dr. Joe Rutledge, Children's Regional Medical Center, Seattle, Washington.)

enzyme hexosaminidase A, which is necessary for the degradation of G_{M2}. More than 100 mutations have been described; most affect protein folding or intracellular transport. The brain is principally affected, because it is most involved in ganglioside metabolism. *The storage of G_{M2}* occurs within neurons, axon cylinders of nerves, and glial cells throughout the CNS. Affected cells appear swollen and sometimes foamy (Fig. 6–9, *A*). Electron microscopy reveals whorled configurations within lysosomes composed of onion-skin layers of membranes (Fig. 6–9, *B*). These pathologic changes are found throughout the CNS (including the spinal cord), peripheral nerves, and autonomic nervous system. The retina usually is involved as well, where the pallor produced by swollen ganglion cells in the peripheral retina results in a contrasting "cherry red" spot in the relatively unaffected central macula.

The molecular basis for neuronal injury is not fully understood. Because in many cases the mutant protein is misfolded, it induces the so-called "unfolded protein" response (Chapter 1). If such misfolded proteins are not stabilized by chaperones, they trigger apoptosis. These findings have spurred clinical trials of *molecular chaperone therapy* for this and similar lysosomal storage diseases. Such therapy involves use of small molecules that increase chaperone synthesis or reduce degradation of misfolded proteins by the proteosomes.

In the most common acute infantile variant of Tay-Sachs disease, infants appear normal at birth, but motor weakness begins at 3 to 6 months of age, followed by neurologic impairment, onset of blindness, and progressively more severe neurologic dysfunctions. Death occurs within 2 or 3 years. Tay-Sachs disease, like other lipidoses, is most common among Ashkenazi Jews, among whom the frequency of heterozygous carriers is estimated to be 1 in 30. Heterozygote carriers can be reliably detected by estimation of the level of hexosaminidase in the serum or by DNA analysis.

Niemann-Pick Disease Types A and B

Type A and type B Niemann-Pick disease are related entities characterized by a primary deficiency of acid sphingomyelinase and the resultant accumulation of sphingomyelin. In type A, characterized by a severe deficiency of sphingomyelinase, the breakdown of sphingomyelin into ceramide and phosphorylcholine is impaired, and excess sphingomyelin accumulates in all phagocytic cells and in the neurons. The macrophages become stuffed with droplets or particles of the complex lipid, imparting a fine vacuolation or foaminess to the cytoplasm (Fig. 6–10). Electron microscopy confirms that the vacuoles are engorged secondary lysosomes that often contain membranous cytoplasmic bodies resembling concentric lamellated myelin figures, sometimes called "zebra" bodies. Because of their high content of phagocytic cells, *the organs most severely affected are the spleen, liver, bone marrow, lymph nodes, and lungs.* The splenic enlargement may be striking. In addition, the entire CNS, including the spinal cord and ganglia, is involved in this tragic, inexorable process. The affected neurons are enlarged and vacuolated as a result of the storage of lipids. This variant manifests itself in infancy with *massive visceromegaly and severe neurologic deterioration.*

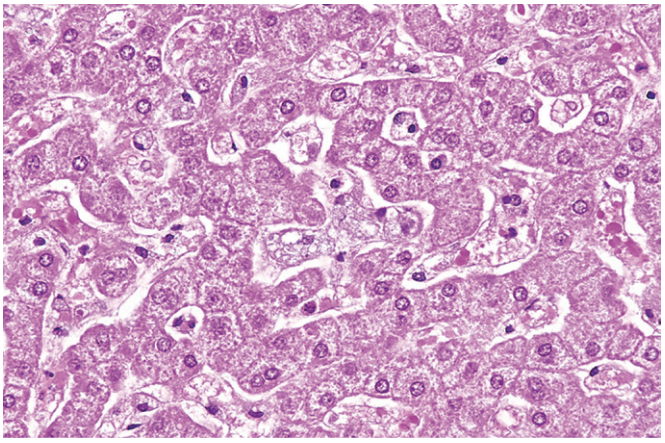


Figure 6–10 Niemann-Pick disease in liver. The hepatocytes and Kupffer cells have a foamy, vacuolated appearance resulting from deposition of lipids.
(Courtesy of Dr. Arthur Weinberg, Department of Pathology, University of Texas Southwestern Medical Center, Dallas, Texas.)

Death usually occurs within the first 3 years of life. By comparison, patients with the type B variant have organomegaly but no neurologic manifestations. Estimation of sphingomyelinase activity in the leukocytes or cultured fibroblasts can be used for diagnosis of suspected cases, as well as for detection of carriers. Antenatal diagnosis is possible by enzyme assays or DNA probe analysis.

Niemann-Pick Disease Type C

Although previously considered to be related to type A and type B Niemann-Pick disease, type C (NPC) is quite distinct at the biochemical and molecular levels and is more common than types A and B combined. Mutations in two related genes, *NPC1* and *NPC2,* can give rise to the disorder, with *NPC1* being responsible for a majority of cases. Unlike most other lysosomal storage diseases, NPC is due to a primary defect in lipid transport. Affected cells accumulate cholesterol as well as gangliosides such as G_{M1} and G_{M2}. Both NPC1 and NPC2 are involved in the transport of free cholesterol from the lysosomes to the cytoplasm. NPC is clinically heterogeneous: The most common form manifests in childhood and is marked by ataxia, vertical supranuclear gaze palsy, dystonia, dysarthria, and psychomotor regression.

Gaucher Disease

Gaucher disease results from mutation in the gene that encodes glucocerebrosidase. There are three autosomal recessive variants of Gaucher disease resulting from distinct allelic mutations. Common to all is variably deficient activity of a glucocerebrosidase that normally cleaves the glucose residue from ceramide. This deficit leads to an accumulation of glucocerebroside, an intermediate in glycolipid metabolism, in the mononuclear phagocytic cells and their transformation into so-called Gaucher cells. Normally the glycolipids derived from the breakdown of senescent blood cells are sequentially degraded by the phagocytic cells of the body particularly in the liver, spleen, and bone marrow. In Gaucher disease, the degradation stops at the level of glucocerebrosides, which accumulate in the phagocytes. These phagocytes—the Gaucher cells—become enlarged, with some reaching a diameter as great as 100 μm, because of the accumulation of distended lysosomes, and acquire a pathognomonic cytoplasmic appearance characterized as "wrinkled tissue paper" (Fig. 6–11). No distinct vacuolation is present. It is evident now that Gaucher disease is caused not just by the burden of storage material but also by activation of the macrophages. High levels of macrophage-derived cytokines, such as interleukins (IL-1, IL-6) and tumor necrosis factor (TNF), are found in affected tissues.

One variant, type I, also called the *chronic nonneuronopathic form,* accounts for 99% of cases of Gaucher disease. It is characterized by clinical or radiographic bone involvement (osteopenia, focal lytic lesions, and osteonecrosis) in 70% to 100% of cases. Additional features are hepatosplenomegaly and the absence of CNS involvement. The spleen often enlarges to massive proportions, filling the entire abdomen. Gaucher cells are found in the liver, spleen, lymph nodes, and bone marrow. Marrow replacement and cortical erosion may produce radiographically visible skeletal lesions, as well as a reduction in the formed elements of blood. Bone changes are believed to be caused by the aforementioned macrophage-derived cytokines. Type I is most common in Ashkenazi Jews; unlike other variants, it is compatible with long life. Types II and III variants are characterized by neurologic signs and symptoms. In type II, these manifestations appear during infancy (*acute infantile neuronopathic form*) and are more severe, whereas in type III, they emerge later and are milder (*chronic neuronopathic form*). Although the liver and spleen also are involved, the clinical features in types II and III are dominated by neurologic disturbances, including convulsions and progressive mental deterioration. The level of glucocerebrosides in leukocytes or cultured fibroblasts is helpful in diagnosis and in the detection of heterozygote carriers.

Current therapy is aimed at lifelong enzyme replacement by infusion of recombinant glucocerebrosidase. A newer form of therapy involves reducing the substrate (glucocerebroside) by oral administration of drugs that inhibit glucocerebroside synthase. Since glucosylceramide is reduced, its accumulation also is reduced. Recent clinical trials in humans have shown considerable promise for this

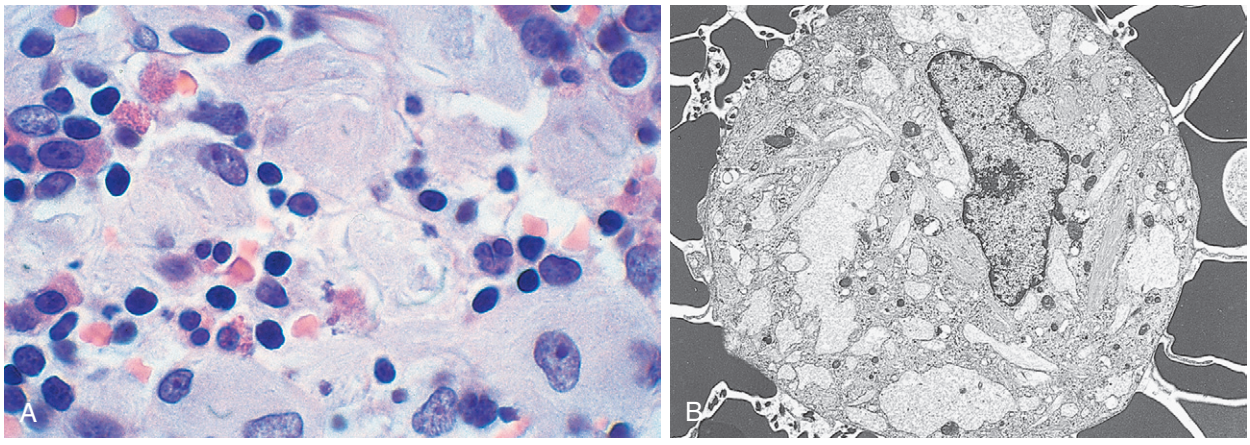


Figure 6–11 Gaucher disease involving the bone marrow. **A,** Gaucher cells with abundant lipid-laden granular cytoplasm. **B,** Electron micrograph of Gaucher cells with elongated distended lysosomes.
(Courtesy of Dr. Matthew Fries, Department of Pathology, University of Texas Southwestern Medical Center, Dallas, Texas.)

modality of therapy, with decrease in splenomegaly and improvements in skeletal disease. On the horizon is glucocerebrosidase gene therapy involving infusion of autologous hematopoietic stem cells transfected with the normal gene.

Mucopolysaccharidoses

Mucopolysaccharidoses (MPSs) are characterized by defective degradation (and therefore excessive storage) of mucopolysaccharides in various tissues. Recall that mucopolysaccharides form a part of ground substance and are synthesized by connective tissue fibroblasts. Most of the mucopolysaccharide is secreted into the ground substance, but a certain fraction is degraded within lysosomes. Multiple enzymes are involved in this catabolic pathway; it is the lack of these enzymes that leads to accumulation of mucopolysaccharides within the lysosomes. Several clinical variants of MPS, classified numerically from MPS I to MPS VII, have been described, each resulting from the deficiency of one specific enzyme. The mucopolysaccharides that accumulate within the tissues include dermatan sulfate, heparan sulfate, keratan sulfate, and (in some cases) chondroitin sulfate.

Hepatosplenomegaly, skeletal deformities, lesions of heart valves, and subendothelial arterial deposits, particularly in the coronary arteries, and lesions in the brain, are common threads that run through all of the MPSs. In many of the more protracted syndromes, coronary subendothelial lesions lead to myocardial ischemia. *Thus, myocardial infarction and cardiac decompensation are important causes of death.* Most cases are associated with *coarse facial features, clouding of the cornea, joint stiffness, and mental retardation.* Urinary excretion of the accumulated mucopolysaccharides often is increased. With all of these disorders except one, the mode of inheritance is autosomal recessive; the exception, Hunter syndrome, is an X-linked recessive disease. Of the seven recognized variants, only two well-characterized syndromes are discussed briefly here.

MPS type I, also known as *Hurler syndrome,* is caused by a deficiency of α-L-iduronidase. In Hurler syndrome, affected children have a life expectancy of 6 to 10 years, and death is often due to cardiac complications. Accumulation of dermatan sulfate and heparan sulfate is seen in cells of the mononuclear phagocyte system, in fibroblasts, and within endothelium and smooth muscle cells of the vascular wall. The affected cells are swollen and have clear cytoplasm, resulting from the accumulation of material positive for periodic acid–Schiff staining within engorged, vacuolated lysosomes. Lysosomal inclusions also are found in neurons, accounting for the mental retardation.

The other well-characterized variant, MPS type II or *Hunter syndrome,* differs from Hurler syndrome in its mode of inheritance (X-linked), the absence of corneal clouding, and often its milder clinical course. As in Hurler syndrome, the accumulated mucopolysaccharides in Hunter syndrome are heparan sulfate and dermatan sulfate, but this results from a deficiency of L-iduronate sulfatase. Despite the difference in enzyme deficiency, an accumulation of identical substrates occurs because breakdown of heparan sulfate and dermatan sulfate requires both α-L-iduronidase and the sulfatase; if either one is missing, further degradation is blocked.

SUMMARY

Lysosomal Storage Diseases

- *Tay-Sachs disease* is caused by an inability to metabolize G_{M2} gangliosides due to lack of the β subunit of lysosomal hexosaminidase. G_{M2} gangliosides accumulate in the CNS and cause severe mental retardation, blindness, motor weakness, and death by 2 to 3 years of age.
- *Niemann-Pick disease types A and B* are caused by a deficiency of sphingomyelinase. In the more severe, type A variant, accumulation of sphingomyelin in the nervous system results in neuronal damage. Lipid also is stored in phagocytes within the liver, spleen, bone marrow, and lymph nodes, causing their enlargement. In type B, neuronal damage is not present.
- *Niemann-Pick disease type C* is caused by a defect in cholesterol transport and resultant accumulation of cholesterol and gangliosides in the nervous system. Affected children exhibit ataxia, dysarthria, and psychomotor regression.
- *Gaucher disease* results from lack of the lysosomal enzyme glucocerebrosidase and accumulation of glucocerebroside in mononuclear phagocytic cells. In the most common, type I variant, affected phagocytes become enlarged (Gaucher cells) and accumulate in liver, spleen, and bone marrow, causing hepatosplenomegaly and bone erosion. Types II and III are characterized by variable neuronal involvement.
- *Mucopolysaccharidoses* result from accumulation of mucopolysaccharides in many tissues including liver, spleen, heart, blood vessels, brain, cornea, and joints. Affected patients in all forms have coarse facial features. Manifestations of Hurler syndrome include corneal clouding, coronary arterial and valvular deposits, and death in childhood. Hunter syndrome is associated with a milder clinical course.

Glycogen Storage Diseases (Glycogenoses)

An inherited deficiency of any one of the enzymes involved in glycogen synthesis or degradation can result in excessive accumulation of glycogen or some abnormal form of glycogen in various tissues. The type of glycogen stored, its intracellular location, and the tissue distribution of the affected cells vary depending on the specific enzyme deficiency. Regardless of the tissue or cells affected, the glycogen most often is stored within the cytoplasm, or sometimes within nuclei. One variant, Pompe disease, is a form of lysosomal storage disease, because the missing enzyme is localized to lysosomes. Most glycogenoses are inherited as autosomal recessive diseases, as is common with "missing enzyme" syndromes.

Approximately a dozen forms of glycogenoses have been described in association with specific enzyme deficiencies. On the basis of pathophysiologic findings, they can be grouped into three categories (Table 6–5):

- *Hepatic type.* Liver contains several enzymes that synthesize glycogen for storage and also break it down into free glucose. Hence, a deficiency of the hepatic enzymes involved in glycogen metabolism is associated with two major clinical effects: *enlargement of the liver due to storage*

Table 6–5 Principal Subgroups of Glycogenoses

Clinicopathologic Category	Specific Type	Enzyme Deficiency	Morphologic Changes	Clinical Features
Hepatic type	Hepatorenal (von Gierke disease, type I)	Glucose-6-phosphatase	*Hepatomegaly*: intracytoplasmic accumulations of glycogen and small amounts of lipid; intranuclear glycogen *Renomegaly*: intracytoplasmic accumulations of glycogen in cortical tubular epithelial cells	In untreated patients, failure to thrive, stunted growth, hepatomegaly, and renomegaly Hypoglycemia due to failure of glucose mobilization, often leading to convulsions Hyperlipidemia and hyperuricemia resulting from deranged glucose metabolism; many patients develop gout and skin xanthomas Bleeding tendency due to platelet dysfunction With treatment (providing continuous source of glucose), most patients survive and develop late complications (e.g., hepatic adenomas)
Myopathic type	McArdle syndrome (type V)	Muscle phosphorylase	*Skeletal muscle only*: accumulations of glycogen predominant in subsarcolemmal location	Painful cramps associated with strenuous exercise Myoglobinuria occurs in 50% of cases Onset in adulthood (>20 yr) Muscular exercise fails to raise lactate level in venous blood Compatible with normal longevity
Miscellaneous type	Generalized glycogenosis (Pompe disease, type II)	Lysosomal glucosidase (acid maltase)	*Mild hepatomegaly*: ballooning of lysosomes with glycogen creating lacy cytoplasmic pattern *Cardiomegaly*: glycogen within sarcoplasm as well as membrane-bound *Skeletal muscle*: similar to heart (see above under cardiomegaly)	Massive cardiomegaly, muscle hypotonia, and cardiorespiratory failure before age 2 Milder adult form with only skeletal muscle involvement manifests with chronic myopathy.

of glycogen and *hypoglycemia due to a failure of glucose production* (Fig. 6–12). *Von Gierke* disease (type I glycogenosis), resulting from a lack of glucose-6-phosphatase, is the most important example of the hepatic form of glycogenosis (Table 6–5).

- *Myopathic type.* In striated muscle, glycogen is an important source of energy. Not surprisingly, most forms of glycogen storage disease affect muscles. When enzymes that are involved in glycolysis are deficient, glycogen storage occurs in muscles and there is an associated muscle weakness due to impaired energy production. Typically, *the myopathic forms of glycogen storage diseases are marked by muscle cramps after exercise, myoglobinuria, and failure of exercise to induce an elevation in blood lactate levels because of a block in glycolysis.* McArdle disease (type V glycogenosis), resulting from a deficiency of muscle phosphorylase, is the prototype of myopathic glycogenoses.
- Type II glycogenosis (*Pompe disease*) is caused by a deficiency of lysosomal acid maltase and so is associated with deposition of glycogen in virtually every organ, but cardiomegaly is most prominent. Most affected patients die within 2 years of onset of cardiorespiratory failure. Therapy with the missing enzyme (glucosidase) can reverse cardiac muscle damage and modestly increase longevity.

SUMMARY

Glycogen Storage Diseases

- Inherited deficiency of enzymes involved in glycogen metabolism can result in storage of normal or abnormal forms of glycogen, predominantly in liver or muscles or in all tissues.
- In the *hepatic form* (von Gierke disease), liver cells store glycogen because of a lack of hepatic glucose-6-phosphatase. There are several *myopathic forms*, including McArdle disease, in which muscle phosphorylase lack gives rise to storage in skeletal muscles and cramps after exercise. In *Pompe disease* there is lack of lysosomal acid maltase, and all organs are affected, but heart involvement is predominant.

Diseases Caused by Mutations in Genes Encoding Proteins That Regulate Cell Growth

As detailed in Chapter 5, two classes of genes, proto-oncogenes and tumor suppressor genes, regulate normal cell growth and differentiation. Mutations affecting these genes, most often in somatic cells, are involved in the

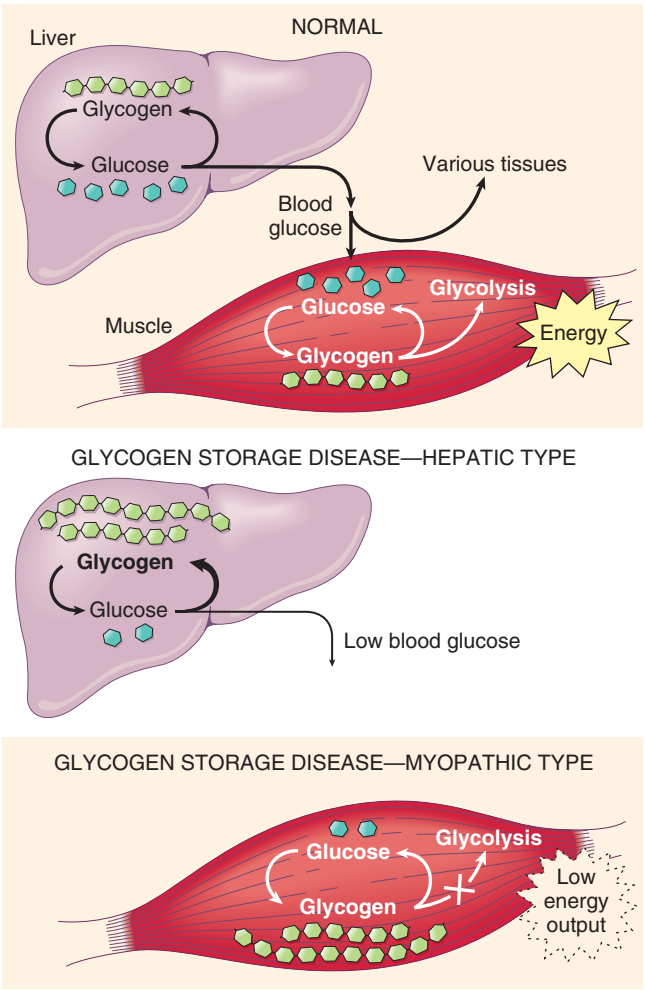


Figure 6–12 Top, A simplified scheme of normal glycogen metabolism in the liver and skeletal muscles. **Middle,** The effects of an inherited deficiency of hepatic enzymes involved in glycogen metabolism. **Bottom,** The consequences of a genetic deficiency in the enzymes that metabolize glycogen in skeletal muscles.

pathogenesis of tumors. In approximately 5% to 10% of all cancers, however, mutations affecting certain tumor suppressor genes are present in all cells of the body, including germ cells and hence can be transmitted to the offspring. These mutant genes predispose the offspring to hereditary tumors, a topic discussed in greater detail in Chapter 5.

COMPLEX MULTIGENIC DISORDERS

Complex multigenic disorders—so-called multifactorial or polygenic disorders—are caused by interactions between variant forms of genes and environmental factors. A genetic variant that has at least two alleles and occurs in at least 1% of the population is called a *polymorphism.* According to the common disease–common variant hypothesis, complex multigenic disorders occur when many polymorphisms, each with a modest effect and low penetrance, are co-inherited. Two additional important facts have emerged from studies of common complex disorders such as type 1 diabetes:

- While complex disorders result from the collective inheritance of many polymorphisms, different polymorphisms vary in significance. For example, of the 20 to 30 genes implicated in type 1 diabetes, 6 or 7 are most important, and a few HLA alleles contribute more than 50% of the risk (Chapter 19).
- Some polymorphisms are common to multiple diseases of the same type, while others are disease-specific. This observation is well illustrated in immune-mediated inflammatory diseases (Chapter 4).

Several normal phenotypic characteristics are governed by multigenic inheritance, such as hair color, eye color, skin color, height, and intelligence. These characteristics (also known as *quantitative trait loci* [QTLs]) show a continuous variation within, as well as across, all population groups. Environmental influences, however, significantly modify the phenotypic expression of complex traits. For example, type 2 diabetes mellitus has many of the features of a complex multigenic disorder. It is well recognized clinically that affected persons often first exhibit clinical manifestations of this disease after weight gain. Thus, obesity as well as other environmental influences, unmasks the diabetic genetic trait. Assigning a disease to this mode of inheritance must be done with caution. Such attribution depends on many factors but first on familial clustering and the exclusion of mendelian and chromosomal modes of transmission. A range of levels of severity of a disease is suggestive of a complex multigenic disorder, but as pointed out earlier, variable expressivity and reduced penetrance of single mutant genes also may account for this phenomenon. Because of these problems, sometimes it is difficult to distinguish between mendelian and multifactorial disorders.

CYTOGENETIC DISORDERS

Chromosomal abnormalities occur much more frequently than is generally appreciated. It is estimated that approximately 1 in 200 newborn infants has some form of chromosomal abnormality. The figure is much higher in fetuses that do not survive to term. It is estimated that in 50% of first-trimester spontaneous abortions, the fetus has a chromosomal abnormality. Cytogenetic disorders may result from alterations in the number or structure of chromosomes and may affect autosomes or sex chromosomes.

Before embarking on a discussion of chromosomal aberrations, it is appropriate to review karyotyping as the basic tool of the cytogeneticist. A *karyotype* is a photographic representation of a stained metaphase spread in which the chromosomes are arranged in order of decreasing length. A variety of techniques for staining chromosomes have been developed. With the widely used Giemsa stain (G banding) technique, each chromosome set can be seen to possess a distinctive pattern of alternating light and dark bands of variable widths (Fig. 6–13). The use of banding techniques allows identification of each chromosome, and can detect and localize structural abnormalities large enough to produce changes in banding pattern (described later).

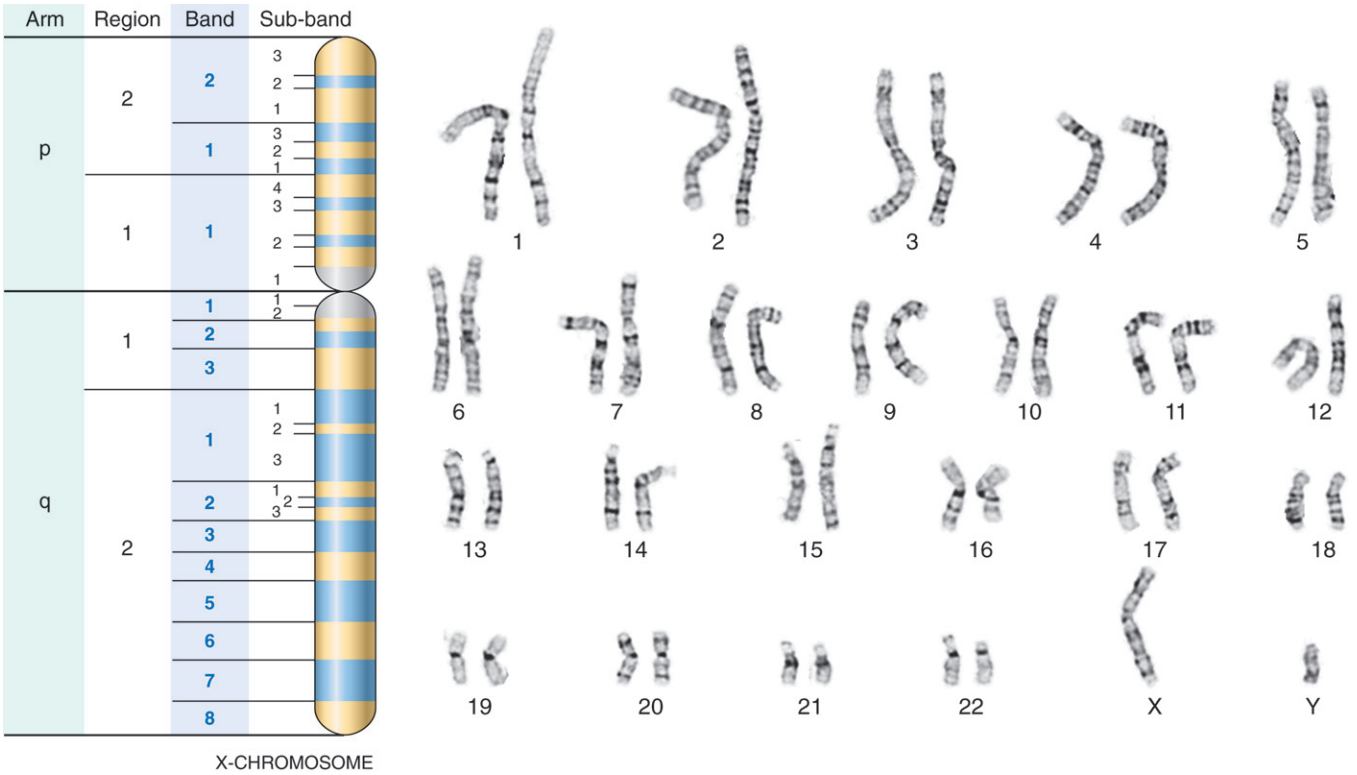


Figure 6–13 G-banded karyotype from a normal male (46,XY). Also shown is the banding pattern of the X-chromosome with nomenclature of arms, regions, bands, and sub-bands.

(Karyotype courtesy of Dr. Stuart Schwartz, Department of Pathology, University of Chicago, Chicago, Illinois.)

Numeric Abnormalities

In humans, the normal chromosome count is 46 (i.e., $2n = 46$). Any exact multiple of the haploid number (n) is called *euploid.* Chromosome numbers such as $3n$ and $4n$ are called *polyploid.* Polyploidy generally results in a spontaneous abortion. Any number that is not an exact multiple of n is called *aneuploid.* The chief cause of aneuploidy is nondisjunction of a homologous pair of chromosomes at the first meiotic division or a failure of sister chromatids to separate during the second meiotic division. The latter also may occur during mitosis in somatic cells, leading to the production of two aneuploid cells. Failure of pairing of homologous chromosomes followed by random assortment (anaphase lag) can also lead to aneuploidy. When nondisjunction occurs at the time of meiosis, the gametes formed have either an extra chromosome ($n + 1$) or one less chromosome ($n - 1$). Fertilization of such gametes by normal gametes would result in two types of zygotes: trisomic, with an extra chromosome ($2n + 1$), or monosomic ($2n - 1$). Monosomy involving an autosome is incompatible with life, whereas trisomies of certain autosomes and monosomy involving sex chromosomes are compatible with life. These, as we shall see, are associated with variable degrees of phenotypic abnormality. *Mosaicism* is a term used to describe the presence of two or more populations of cells with different complements of chromosomes in the same individual. In the context of chromosome numbers, postzygotic mitotic nondisjunction would result in the production of a trisomic and a monosomic daughter cell; the descendants of these cells would then produce a mosaic. As discussed later, mosaicism affecting sex chromosomes is common, whereas autosomal mosaicism is not.

Structural Abnormalities

Structural changes in the chromosomes usually result from chromosomal breakage followed by loss or rearrangement of material. Such changes usually are designated using a cytogenetic shorthand in which *p* (French, *petit*) denotes the short arm of a chromosome, and *q,* the long arm. Each arm is then divided into numbered regions (1, 2, 3, and so on) from centromere outward, and within each region the bands are numerically ordered (Fig. 6–13). Thus, 2q34 indicates chromosome 2, long arm, region 3, band 4. The patterns of chromosomal rearrangement after breakage (Fig. 6–14) are as follows:

- *Translocation* implies transfer of a part of one chromosome to another chromosome. The process is usually reciprocal (i.e., fragments are exchanged between two chromosomes). In genetic shorthand, translocations are indicated by *t* followed by the involved chromosomes in numeric order—for example, 46,XX,t(2;5)(q31;p14). This notation would indicate a reciprocal translocation involving the long arm (q) of chromosome 2 at region 3, band 1, and the short arm of chromosome 5, region 1, band 4. When the entire broken fragments are exchanged, the resulting balanced reciprocal translocation (Fig. 6–14) is not harmful to the carrier, who has the normal number of chromosomes and the full complement of genetic material. However, during gametogenesis,

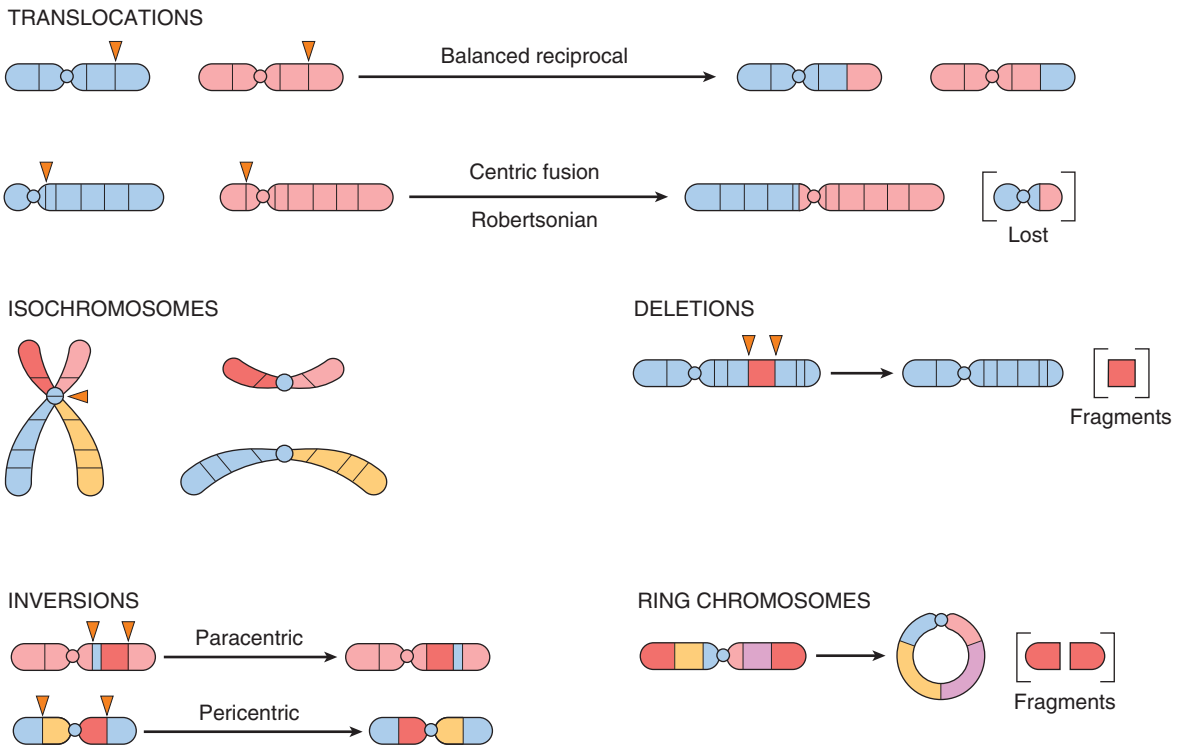


Figure 6–14 Types of chromosomal rearrangements.

abnormal (unbalanced) gametes are formed, resulting in abnormal zygotes. A special pattern of translocation involving two acrocentric chromosomes is called *centric fusion type,* or *robertsonian,* translocation. The breaks typically occur close to the centromere, affecting the short arms of both chromosomes. Transfer of the segments leads to one very large chromosome and one extremely small one (Fig. 6–14). The short fragments are lost, and the carrier has 45 chromosomes. Because the short arms of all acrocentric chromosomes carry highly redundant genes (e.g., ribosomal RNA genes), such loss is compatible with survival. However, difficulties arise during gametogenesis, resulting in the formation of unbalanced gametes that could lead to abnormal offspring.

- *Isochromosomes* result when the centromere divides horizontally rather than vertically. One of the two arms of the chromosome is then lost, and the remaining arm is duplicated, resulting in a chromosome with two short arms only or two long arms only. The most common isochromosome present in live births involves the long arm of the X chromosome and is designated *i(Xq)*. When fertilization occurs by a gamete that contains a normal X chromosome, the result is monosomy for genes on Xp and trisomy for genes on Xq.
- *Deletion* involves loss of a portion of a chromosome. A single break may delete a terminal segment. Two interstitial breaks, with reunion of the proximal and distal segments, may result in loss of an intermediate segment. The isolated fragment, which lacks a centromere, almost never survives, and thus many genes are lost.
- *Inversions* occur when there are two interstitial breaks in a chromosome, and the segment reunites after a complete turnaround.
- A *ring chromosome* is a variant of a deletion. After loss of segments from each end of the chromosome, the arms unite to form a ring.

General Features of Chromosomal Disorders

- Chromosomal disorders may be associated with absence (deletion, monosomy), excess (trisomy), or abnormal rearrangements (translocations) of chromosomes.
- In general, loss of chromosomal material produces more severe defects than does gain of chromosomal material.
- Excess chromosomal material may result from a complete chromosome (as in trisomy) or from part of a chromosome (as in robertsonian translocation).
- Imbalances of sex chromosomes (excess or loss) are tolerated much better than are similar imbalances of autosomes.
- Sex chromosomal disorders often produce subtle abnormalities, sometimes not detected at birth. Infertility, a common manifestation, cannot be diagnosed until adolescence.
- In most cases, chromosomal disorders result from de novo changes (i.e., parents are normal, and risk of recurrence in siblings is low). An uncommon but important exception to this principle is exhibited by the translocation form of Down syndrome (described later).